UNIVERSITY OF CALIFORNIA SAN DIEGO

Characterizing Unstructured Motor Behaviors in the Epilepsy Monitoring Unit

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Electrical Engineering
(Medical Devices & Systems)

by

Paolo Gutierrez Gabriel

Committee in charge:

Professor Vikash Gilja, Chair
Professor Eric Halgren
Professor Duygu Kuzum
Professor Cory Miller
Professor Truong Nguyen
Professor Sonya Wang

2019

The dissertation of Paolo Gutierrez Gabriel is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

_____

Chair

University of California San Diego

2019

EPIGRAPH

You can't ever reach perfection, but you can believe in an asymptote toward which you are ceaselessly striving.

Paul Kalanithi, "When Breath Becomes Air"

I slept and dreamt that life was joy. I awoke and saw that life was service. I acted and behold, service was joy.

Rabindranath Tagore

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Vikash Gilja, for his guidance, support, and inspiration. I am indebted to him for the countless opportunities he provided, and for nurturing me into a better researcher. It has been an honor and a privilege to have witnessed the growth of the lab alongside you. My parents have provided me with the means and the encouragement to pursue my goals, no matter where they take me. I am thankful for their unconditional love and support throughout my life, particularly over the past 6 years. It has been a blessing to connect with so many people during my time in San Diego. My acknowledgements to them are short - to some I say 'Thank you', to others I say 'You are welcome', and to the last of us I say 'I love and support you'.

The chapters of this dissertation consist of published and submitted journal articles. The dissertation author was either the primary or secondary investigator and author of each of these papers.

- Chapter 2 is a reprint of the material as it appears in Gabriel, Paolo, Werner K. Doyle, Orrin Devinsky, Daniel Friedman, Thomas Thesen, and Vikash Gilja. "Neural correlates to automatic behavior estimations from rgb-d video in epilepsy unit." In *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference*, pp. 3402-3405. IEEE, 2016.

- Chapter 3 is a reprint of the material as it appears in Alasfour, Abdulwahab, Paolo Gabriel, Xi Jiang, Isaac Shamie, Lucia Melloni, Thomas Thesen, Patricia Dugan, Daniel Friedman, Werner Doyle, Orin Devinsky, David Gonda, Shifteh Sattar, Sonya Wang, Eric Halgren, and Vikash Gilja. "Coarse behavioral context decoding." *Journal of neural engineering* 16, no. 1 (2019): 016021.

- Chapter 4 is a reprint of the material as it appears in Gabriel, Paolo Gutierrez, Kenny Chen, Abdulwahab Alasfour, Tejaswy Pailla, Werner Doyle, Orrin Devinsky, Daniel Friedman, Patricia Dugan, Lucia Melloni, Thomas Thesen, David Gonda, Shifteh Sattar, Sonya Wang, and Vikash Gilja. "Neural correlates of unstructured motor behaviors." *Journal of neural engineering* (2019).

- Chapter 5 is a reprint of the material as it appears in Chen, Kenny, Paolo Gabriel, Abdulwahab Alasfour, Chenghao Gong, Werner K. Doyle, Orrin Devinsky, Daniel Friedman, Patricia Dugan, Lucia Melloni, Thomas Thesen, David Gonda, Shifteh Sattar, Sonya Wang, and Vikash Gilja. "Patient-specific pose estimation in clinical environments." *IEEE Journal of Translational Engineering in Health and Medicine* 6 (2018): 1-11.

VITA

2013	Bachelor of Science in Engineering Physics, Stanford University, Stanford
2013 – 2019	Graduate Student Researcher, University of California, San Diego
2015 , 2018	Teaching Assistant, University of California, San Diego
2015	Master of Science in Electrical Engineering (Medical Devices & Systems), University of California, San Diego
2019	Doctor of Philosophy in Electrical Engineering (Medical Devices & Systems), University of California, San Diego
2019 –	Computer Vision Engineer, LookDeep Health, Oakland, CA

PUBLICATIONS

Gabriel, Paolo, Werner K. Doyle, Orrin Devinsky, Daniel Friedman, Thomas Thesen, and Vikash Gilja. "Neural correlates to automatic behavior estimations from rgb-d video in epilepsy unit." In *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference*, pp. 3402-3405. IEEE, 2016.

Chen, Kenny, **Paolo Gabriel**, Abdulwahab Alasfour, Chenghao Gong, Werner K. Doyle, Orrin Devinsky, Daniel Friedman et al."Patient-specific pose estimation in clinical environments." *IEEE Journal of Translational Engineering in Health and Medicine* 6 (2018): 1-11.

Alasfour, Abdulwahab, **Paolo Gabriel**, Xi Jiang, Isaac Shamie, Lucia Melloni, Thomas Thesen, Patricia Dugan et al. "Coarse behavioral context decoding." *Journal of neural engineering* 16, no. 1 (2019): 016021.

Gabriel, Paolo Gutierrez, Kenny Chen, Abdulwahab Alasfour, Tejaswy Pailla, Werner Doyle, Orrin Devinsky, Daniel Friedman et al. "Neural correlates of unstructured motor behaviors." *Journal of neural engineering* (2019).

Martin, Joel, **Paolo Gutierrez Gabriel**, Jeffrey Gold, Richard Haas, Sue Davis, David Gonda, Cia Sharpe et al. "Optical flow estimation improves automated seizure detection in neonatal EEG." *Journal of Clinical Neurophysiology* (2020, in review).

ABSTRACT OF THE DISSERTATION

Characterizing Unstructured Motor Behaviors
in the Epilepsy Monitoring Unit

by

Paolo Gutierrez Gabriel

Doctor of Philosophy in Electrical Engineering
(Medical Devices & Systems)

University of California San Diego, 2019

Professor Vikash Gilja, Chair

Key advancements in recording hardware, data computation, clinical care, and cognitive science continue to drive new possibilities in how humans and machines can interact directly through thought. Neural data analyses with these advancements has progressed neuroscience research in functional brain mapping and brain-computer interfaces (BCIs). Much of our knowledge about BCIs is informed by data collected through carefully controlled experiments. Constraining BCI experiments with structured paradigms allows researchers to collect a high number of consistent data in a short amount of time, while also controlling for external confounds. Very little is currently known about how well these task-based relationships extend to daily life, in part because collecting data outside of the lab is challenging. To further understand natural brain activity, we must study more complex behaviors in more environmentally relevant settings. The results of this dissertation address three general challenges to studying neural correlates to unstructured behaviors. First, we continuously monitored unstructured human movements in the epilepsy monitoring unit using a video sensor synchronized to clinical intracortical electrodes. Second, we annotated unstructured behaviors

from these video using both manual and computer vision methods. Finally, analyzed neural features with respect to unstructured human movements, and evaluated the performance of features identified in previous task-based studies. The preliminary nature of this work means that a majority of our demonstrations are whether the continuous paradigm can be leveraged, how one might go about leveraging it, and evaluations that tie our results back to earlier task-based studies. Our advances here motivate future works that focus more intently on what types of behaviors and neural signal features to explore.

Chapter 1

Introduction

1.1 Background

Key advancements in recording hardware, data computation, clinical care, and cognitive science continue to drive new possibilities in how humans and machines can interact directly through thought. Advances in medical devices and systems allow for recording and storing brain activity with increasingly higher temporal and spatial resolutions [1, 2]. Both invasive and non-invasive methods are available to monitor various modes of signals, such as electrical potentials. Recording such signals progress our understanding of the brain and its functions. Medical advances provide the means to interface these technologies with humans, as needed. For example, invasive brain recording methods such as electrocorticography (ECoG) and stereotaxic electroencephalograph (sEEG) require neurosurgery and clinical care while undergoing monitoring. Clinically, such methods are important for localizing impacted brain regions in humans who are neurologically impaired [3–5]. In the case of a person with epilepsy, clinicians can monitor these signals to make an informed decisions about the specific type of treatment is needed to stop seizing. Finally, many experiments are motivated by a greater understanding of neural signal properties [6, 7] and cortical function [8, 9]. Particularly, studies like [10] have identified spectral responses to various movement modalities in the high-frequency band (76–100Hz, HFB) and low-frequency band (8–32Hz, LFB) of cortical signals near central sulcus. Because the coverage provided by ECoG and sEEG broadly covers the brain, they enable analysis of behaviors ranging from task-related responses to abstract contexts.

Since the first successful decoding of neural signals to control a robotic manipulator [11], researchers have pushed human brain-computer interfaces (BCIs) to additional applications, including functional clinical mapping [12–14], speech decoding [15–17], and motor prosthetic control [18–23]. Much of our knowledge about BCIs is informed by data collected through carefully controlled experiments in laboratory or clinical environments. Constraining BCI experiments with highly structured paradigms and settings

allows researchers to collect a high number of consistent trials for a given behavior, while controlling for external confounding factors, when the timing of recording sessions is limited to several hours or less. By reducing behavioral complexity and variability, researchers have identified specific relationships between the brain and functional behaviors like movement [24] and speech [25]. However, very little is currently known about how well such these task-based relationships extend to daily life [26], in part because collecting data outside of the lab is challenging. To further understand natural brain activity, we must study more complex behaviors in more environmentally relevant settings.

1.2 Objective

BCI studies show that humans can control computer software and prostheses with brain signals for very specific applications, for example [27–29], and that BCI based tasks can be utilized in the clinic for functional brain mapping [12–14]. The generalization of BCIs across a wider set of behavioral and environmental contexts is largely unstudied. For example, it remains unanswered as to how the human brain controls movements not evoked by tasks. Studies like [30] show neural spiking activity occurring during self-paced motor behavior not induced by tasks that is similar to neural spiking activity induced by a task. But what is unanswered is how unstructured, daily-life motor behaviors compare with specific controlled motor output.

In this work, continuous multi-modal datasets were collected in the EMU from subjects with electrodes implanted in and on their brain. Hours of video of each subject were recorded alongside neural signals from electrodes across many regions of human cortex. Unlike traditional task-based studies that target highly repeatable and instructed behaviors, these data were collected under naturalistic conditions (*e.g.* subject is stretching, eating, playing), thus providing a richer collection of unstructured human behaviors.

We hypothesize that this continuous monitoring paradigm will provide examples of complex and nuanced movement behaviors that complement task-evoked movements. Studying neural activity with respect to these more natural movements may inform future BCIs, whereas the movement annotation tools developed for this study may augment behavior monitoring in the clinic. To demonstrate the feasibility of such an approach, this work follows two paths: 1) semi-automated, quantitative tracking and trialization of subject movement from video using available computer vision and signal processing techniques; and 2) application of discrete classification techniques on neural signal features previously associated with instructed movements. This work serves as a preliminary demonstration that simultaneously recording intracortical activity and unstructured movements can serve as an additional means to assess neural correlates to human

behavior.

1.2.1 Challenges

One of the greatest challenges to studying neural activity in unstructured settings is properly curating usable data. Typically, subjects in the EMU are surrounded by medical equipment and clinical staff in an ever-shifting environment. Any of these factors may inject unwanted noise or artifact into either the video or neural recording; task-based studies are typically designed to limit impact of these factors. In addition to these challenges, frequency of activity varies from subject to subject, and periods of movement are relatively sparse compared to periods of rest. Identifying these critical periods while avoiding confounding periods requires careful inspection and annotation. On a small scale, this is usually done by hand, but the task grows in difficulty as longer datasets are recorded. Longer datasets are necessary to capture more examples of behavior and to document a wider range of behaviors. While there are methods to automate certain annotations [31, 32], few are sufficiently reliable in noisy environments. By addressing these challenges, however, one may be able to study neural activity in unstructured settings.

1.3 Thesis Overview

The results of this dissertation address three general challenges to studying neural correlates to unstructured behaviors:

- Continuous monitoring of unstructured human movements in the epilepsy monitoring unit
- Annotating unstructured behaviors from clinical video
- Analyzing neural features with respect to unstructured human movements

The remainder of this dissertation consists of four papers published on these topics, organized as follows:

In Chapter 2, we introduce our data acquisition system alongside a preliminary demonstration showing that neural activity from the continuous monitoring paradigm can be used to map electrode response to left- and right-sided movements. The main methods in this work include neural signal processing of ECoG, hand tracking using edge-based image features, and functional mapping using non-parametric statistical tests. Chapter 2 is a reprint of the material as it appears in Gabriel, Paolo, Werner K. Doyle, Orrin Devinsky, Daniel Friedman, Thomas Thesen, and Vikash Gilja. “Neural correlates to automatic behavior estimations

from rgb-d video in epilepsy unit." In *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference*, pp. 3402-3405. IEEE, 2016.

In Chapter 3, we explore neural correlates to coarsely labeled behaviors. These labels describe abstract behavior contexts such as 'Resting' and 'Dialogue', as described by human annotators using a rubric while watching 5-minute video segments of video recorded in the EMU. The main methods in this work describe neural signal processing of ECoG and sEEG, and rigorous evaluation of context classification using binned spectral power as features. Chapter 3 is a reprint of the material as it appears in Alasfour, Abdulwahab, Paolo Gabriel, Xi Jiang, Isaac Shamie, Lucia Melloni, Thomas Thesen, Patricia Dugan et al. "Coarse behavioral context decoding." *Journal of neural engineering* 16, no. 1 (2019): 016021.

In Chapter 4, we demonstrate an evaluation framework to study neural activity related to natural movements not evoked by a task, annotated over hours of video. We found that neural signal features (8–32 Hz and 76–100 Hz power) previously identified from task-based experiments are also modulated before and during a variety of movement behaviors. These movement behaviors are coarsely labeled by time period and movement side (e.g. 'Idle' and 'Move', 'Right' and 'Left'); movements within a label can include a wide variety of uninstructed behaviors. A rigorous nested cross-validation framework was used to classify both movement onset and lateralization with statistical significance for all subjects. Chapter 4 is a reprint of the material as it appears in Gabriel, Paolo Gutierrez, Kenny Chen, Abdulwahab Alasfour, Tejaswy Pailla, Werner Doyle, Orrin Devinsky, Daniel Friedman et al. "Neural correlates of unstructured motor behaviors." *Journal of neural engineering* (2019).

In Chapter 5, we present a framework for adapting existing pose estimation models to improve their robustness in the clinical environment. The extended system yields more consistent and accurate posture annotations when compared to two state-of-the-art generalized pose tracking algorithms for three hospital patients recorded in two research clinics. The proposed framework trains subject-specific convolutional neural network (CNN) models on a subset of a patient-specific RGB video recording chosen to maximize the feature variance of each joint. Scene lighting changes are also accommodated. Finally, these estimations are refined over time through a Kalman filter with fitted noise parameters. Chapter 5 is a reprint of the material as it appears in Chen, Kenny, Paolo Gabriel, Abdulwahab Alasfour, Chenghao Gong, Werner K. Doyle, Orrin Devinsky, Daniel Friedman et al."Patient-specific pose estimation in clinical environments." *IEEE Journal of Translational Engineering in Health and Medicine* 6 (2018): 1-11.

References

[1] Jonathan Viventi, Dae-Hyeong Kim, Leif Vigeland, Eric S Frechette, Justin A Blanco, Yun-Soung Kim, Andrew E Avrin, Vineet R Tiruvadi, Suk-Won Hwang, Ann C Vanleer, et al. Flexible, foldable, actively multiplexed, high-density electrode array for mapping brain activity in vivo. *Nature neuroscience*, 14(12):1599, 2011.

[2] Elon Musk et al. An integrated brain-machine interface platform with thousands of channels. *BioRxiv*, page 703801, 2019.

[3] Scott B Wilson, Christine A Turner, Ronald G Emerson, and Mark L Scheuer. Spike detection ii: automatic, perception-based detection and clustering. *Clinical neurophysiology*, 110(3):404–411, 1999.

[4] Mark L Scheuer and Scott B Wilson. Data analysis for continuous eeg monitoring in the icu: seeing the forest and the trees. *Journal of clinical neurophysiology*, 21(5):353–378, 2004.

[5] F Fürbass, P Ossenblok, M Hartmann, H Perko, AM Skupch, G Lindinger, L Elezi, E Pataraia, AJ Colon, C Baumgartner, et al. Prospective multi-center study of an automatic online seizure detection system for epilepsy monitoring units. *Clinical Neurophysiology*, 126(6):1124–1131, 2015.

[6] György Buzsáki, Costas A Anastassiou, and Christof Koch. The origin of extracellular fields and currents—eeg, ecog, lfp and spikes. *Nature reviews neuroscience*, 13(6):407, 2012.

[7] Scott R Cole and Bradley Voytek. Brain oscillations and the importance of waveform shape. *Trends in Cognitive Sciences*, 21(2):137–149, 2017.

[8] K Hara, S Uematsu, R Lesser, B Gordon, J Hart, and E Vining. Representation of primary motor cortex in humans: studied with chronic subdural grid. *Epilepsia*, 32(SUPPL.):23–24, 1991.

[9] Robert Prueckl, Christoph Kapeller, Cristhian Potes, Milena Korostenskaja, Gerwin Schalk, Ki H Lee, and Christoph Guger. cortiq-clinical software for electrocorticographic real-time functional mapping of the eloquent cortex. In *Conf Proc IEEE Eng Med Biol Soc*, volume 2013, pages 6365–6368, 2013.

[10] Kai J Miller, Eric C Leuthardt, Gerwin Schalk, Rajesh PN Rao, Nicholas R Anderson, Daniel W Moran, John W Miller, and Jeffrey G Ojemann. Spectral changes in cortical surface potentials during motor movement. *Journal of Neuroscience*, 27(9):2424–2432, 2007.

[11] Fetz E. Operant conditioning of cortical unit activity. *Science*, **163**(3870):955–958, 1969.

[12] Gerwin Schalk, Eric C Leuthardt, Peter Brunner, Jeffrey G Ojemann, Lester A Gerhardt, and Jonathan R Wolpaw. Real-time detection of event-related brain activity. *Neuroimage*, 43(2):245–249, 2008.

[13] Korostenskaja M, Wilson A, Rose D, Brunner P, Schalk G, Leach J, Mangano F, Fujiwara H, Rozhkov L, Harris E, et al. Real-time functional mapping with electrocorticography in pediatric epilepsy: comparison with fmri and esm findings. *Clinical EEG and Neuroscience*, **45**(3):205–211, 2014.

[14] Peter Brunner, Anthony L Ritaccio, Timothy M Lynch, Joseph F Emrich, J Adam Wilson, Justin C Williams, Erik J Aarnoutse, Nick F Ramsey, Eric C Leuthardt, Horst Bischof, et al. A practical procedure for real-time functional mapping of eloquent cortex using electrocorticographic signals in humans. *Epilepsy & Behavior*, 15(3):278–286, 2009.

[15] Bouchard K, Mesgarani N, Johnson K, and Chang E. Functional organization of human sensorimotor cortex for speech articulation. *Nature*, **495**(7441):327–332, March 2013.

[16] Mesgarani N and Chang E. Selective cortical representation of attended speaker in multi-talker speech perception. *Nature*, **485**(7397):233–6, May 2012.

[17] Miguel Angrick, Christian Herff, Emily Mugler, Matthew C Tate, Marc W Slutzky, Dean J Krusienski, and Tanja Schultz. Speech synthesis from ecog using densely connected 3d convolutional neural networks. *bioRxiv*, page 478644, 2018.

[18] Wang W, Collinger J, Degenhart A, Tyler-Kabara E, Schwartz A, Moran D, Weber D, Wodlinger B, Vinjamuri R, Ashmore R, Kelly J, and Boninger M. An Electrocorticographic Brain Interface in an Individual with Tetraplegia. *PLoS ONE*, **8**, 2013.

[19] Carmena J, Lebedev M, Crist R, O'Doherty J, Santucci D, Dimitrov D, Patil P, Henriquez C, and Nicolelis M. Learning to control a brain–machine interface for reaching and grasping by primates. *PLoS Biology*, **1**(2), 10 2003.

[20] Chaplin J, Moxon K, Markowitz R, and Nicolelis M. Real-time control of a robot arm using simultaneously recorded neurons in the motor cortex. *Nature*, **2**:664–670, 1999.

[21] Gilja V, Nuyujukian P, Chestek C, Cunningham J, Krishna S, et al. A high-performance neural prosthesis enabled by control algorithm design. *Nature Neuroscience*, **15**:1752–1758, 2012.

[22] Hochberg L, Serruya M, Friehs G, Mukand J, Saleh M, Caplan A, Branner A, Chen D, Penn R, and Donoghue J. Neuronal ensemble control of prosthetic devices by a human with tetraplegia. *Nature*, **442**:164–171, 2006.

[23] Velliste M, Perel S, Spalding C, Whitford A, and Schwartz A. Cortical control of a prosthetic arm for self-feeding. *Nature*, **458**:1098–1101, 2008.

[24] Kai J Miller, Christopher J Honey, Dora Hermes, Rajesh PN Rao, Jeffrey G Ojemann, et al. Broadband changes in the cortical surface potential track activation of functionally diverse neuronal populations. *Neuroimage*, 85:711–720, 2014.

[25] Spencer Kellis, Kai Miller, Kyle Thomson, Richard Brown, Paul House, and Bradley Greger. Decoding spoken words using local field potentials recorded from the cortical surface. *Journal of neural engineering*, 7(5):056007, 2010.

[26] Dastjerdi M, Ozker M, Foster B, Rangarajan V, and Parvizi J. Numerical processing in the human parietal cortex during experimental and natural conditions. *Nature Communications*, **4**:2528, January 2013.

[27] Anna Buttfield, Pierre W Ferrez, and Jd R Millan. Towards a robust bci: error potentials and online learning. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 14(2):164–168, 2006.

[28] Leigh R Hochberg, Daniel Bacher, Beata Jarosiewicz, Nicolas Y Masse, John D Simeral, Joern Vogel, Sami Haddadin, Jie Liu, Sydney S Cash, Patrick van der Smagt, et al. Reach and grasp by people with tetraplegia using a neurally controlled robotic arm. *Nature*, 485(7398):372, 2012.

[29] Inchul Choi, Ilsun Rhiu, Yushin Lee, Myung Hwan Yun, and Chang S Nam. A systematic review of hybrid brain-computer interfaces: Taxonomy and usability perspectives. *PloS one*, 12(4):e0176674, 2017.

[30] Mariska J Vansteensel, Elmar GM Pels, Martin G Bleichner, Mariana P Branco, Timothy Denison, Zachary V Freudenburg, Peter Gosselaar, Sacha Leinders, Thomas H Ottens, Max A Van Den Boom, et al. Fully implanted brain–computer interface in a locked-in patient with als. *New England Journal of Medicine*, 375(21):2060–2066, 2016.

[31] Jamie Shotton, Toby Sharp, Alex Kipman, Andrew Fitzgibbon, Mark Finocchio, Andrew Blake, Mat Cook, and Richard Moore. Real-time human pose recognition in parts from single depth images. *Communications of the ACM*, 56(1):116–124, 2013.

[32] T. Pfister. *Advancing Human Pose and Gesture Recognition*. PhD thesis, University of Oxford, 2015.

Chapter 2

Neural Correlates to Automatic Behavior Estimations from RGB-D Video in Epilepsy Unit

Abstract

To augment neural monitoring, a minimally intrusive multi-modal capture system was designed and implemented in the epilepsy clinic. This system provides RGB-D audio-video synchronized with patient electrocorticography (ECoG), which records neural activity across cortex. We propose an automated approach to studying the human brain in a naturalistic setting. We demonstrate coarse functional mapping of ECoG electrodes correlated to contralateral arm movements. Motor electrode mapping was generated by analyzing continuous movement data recorded over several hours from epilepsy patients in hospital rooms. From these recordings we estimate the kinematics of patient hand movement behaviors using computer vision algorithms. We compare movement behaviors to neural data collected from ECoG, specifically high–γ (70–110 Hz) spectral features. We present a functional map of electrode responses to natural arm movements, generated using a statistical test. We demonstrate that our approach has the potential to aid in the development of automated functional brain mapping using continuous video and neural recordings of patients in clinical settings.

2.1 Introduction

Detailed recordings of synchronous brain activity and behavior have the potential to improve identification of eloquent brain regions, informing resective brain surgery for individuals diagnosed with brain tumors and epilepsy. Functional brain mappings from such recordings rely upon the interpretation of changes in recorded brain signals. Recent clinical studies demonstrate that these interpretations are possible within many behavioral, sensory, and cognitive domains (e.g. motor [1], speech [2, 3], and number abstractions [4]). Such studies are often constrained and thus informed by data collected through short and highly constructed tasks. Patients are also impacted by surgery and pharmaceutical interventions, which limit their participation in task-driven studies [5]. Strict use of structured tasks provide ground-truth labels for the study, but leave the relationship between neural activity and natural human behaviors largely unexplored [6, 7].

One proposed strategy for mapping brain signals is to leverage external monitoring of naturalistic behavior and relate it to neural activity. Long-term recordings of patient activity throughout the day provide detailed accounts of complex behavior not typically found in task-driven studies, such as reading and eating. [8] demonstrates unsupervised decoding of coarse naturalistic human behaviors using automated video and audio annotations from clinical recordings. This work suggests that one can extract information related to behavioral states from continuously recorded electrocorticography (ECoG). We pursue this problem at a finer resolution of time and behavior by designing and deploying a minimally intrusive multi-modal capture system.

We collect multi-day data sets rich with natural human behaviors alongside high-quality neural signals from electrodes implanted across many regions of human cortex. Specifically, we are examining spatial and temporal neural features that predict behaviors of interest. A color and depth (RGB-D) recording system that synchronizes with standard clinical ECoG monitors has been deployed in order to facilitate neural and behavioral analyses on a timescale of 10's of milliseconds. Our work pursues two paths, as outlined in Fig. 2.1: one is automated, quantitative behavioral tracking of synchronized RGB-D video using available computer vision and signal processing algorithms; the other is the application of machine learning and statistical techniques to extract predictive neural features. In joining the two paths with coarse statistical analysis, we demonstrate an approach to mapping cortical activity with contralateral arm movements.

2.2 Data Capture

2.2.1 Subjects

Two patients (NY531, NY540) with intractable epilepsy underwent invasive monitoring to localize epileptogenic zones before surgical resection. During their stay at the New York University (NYU) Langone Epilepsy Unit in New York, patients gave their informed consent and were enrolled in the study. The study was approved by the Internal Review Board of NYU.

2.2.2 Recordings

The subjects were implanted with platinum electrode arrays configured as a combination of strips, numbering between 8 and 10 in total, and a single 'grid' (8x8). The electrodes were embedded in SILASTIC sheets (2.3 mm exposed diameter, 10 mm center-to-center spacing) [9], and placed directly on the (left, right) cortical surface. For each patient, over 100 clinical electrode contacts were implanted across multiple brain regions, as demonstrated for NY531 in Fig. 2a. The placement of electrodes was based entirely on clinical considerations for the identification of seizure foci. ECoG activity was recorded from 0.1 to 230 Hz using Nicolet clinical amplifiers, and the data was digitized and sampled at 512 Hz while referenced to a screw bolted to the skull.

A custom mounting unit was designed to allow for minimally intrusive RGB-D recording using a Microsoft Kinect v2 sensor. This commercial device allows for simultaneous audio, video, and depth recording. These data streams are sampled at 16kHz, 30Hz, and 30Hz, respectively. The recording system

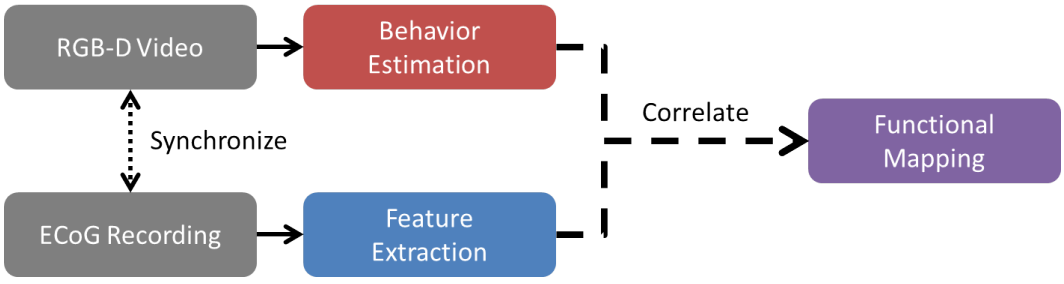


Figure 2.1 **Overview of the full research pipeline for functional brain mapping**. First, electrocorticography (ECoG) recordings were aligned to the RGB-D streams by matching a periodic random-valued signal sent to both systems. The streams were aligned within 25-ms of each other. Using computer vision algorithms, the speed of left and right hands were tracked and averaged into 200ms bins. Aligned ECoG recordings were band-pass filtered into high–γ (70-110 Hz) band. The analytic amplitude for each channel was then calculated using the Hilbert transform, and the resulting signals were binned into 200ms, non-overlapping windows. The distributions of spectral features for each electrode between contralateral movements and "non-movements" were compared using a rank-sum test.

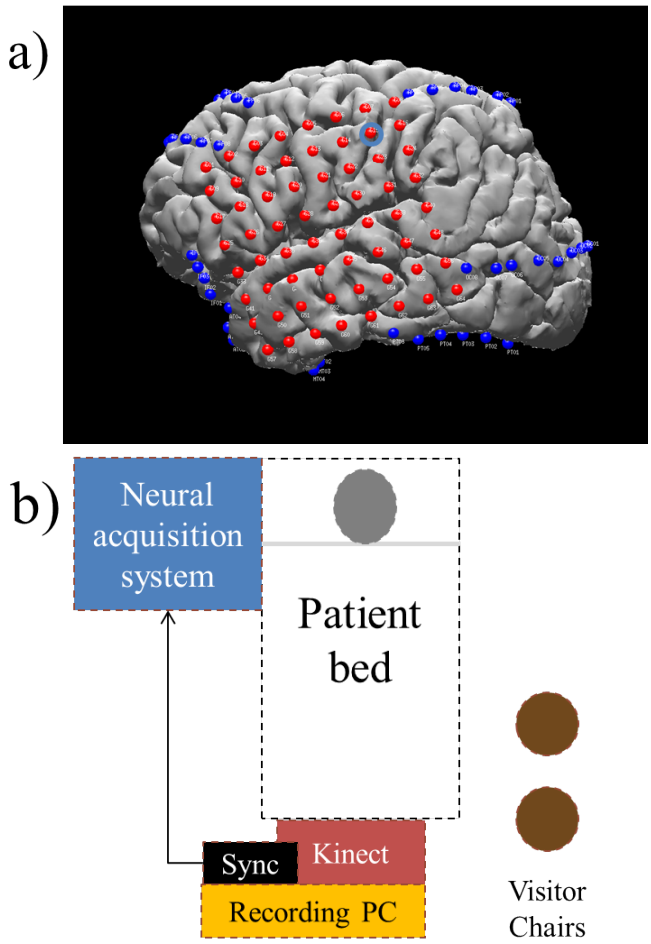


Figure 2.2 a) ECoG electrode map reconstruction of NY531. For each patient, over 100 clinical electrode contacts were implanted across multiple brain regions, and their placement was based entirely on clinical considerations for identifying seizure foci. b) Ideal setup for data collection in the clinic. Data is collected by a system consisting of the Kinect, a local recording computer, and Arduino synchronizer. The system allows for simultaneous audio, video, and depth recording. The sensor can be quickly deployed onto the foot-board of the hospital bed, allowing for an unobstructed bedside view of the patient and their immediate surroundings.

consists of the sensor, a PC for recording and storage, and a custom designed synchronizer connected to the patient's neural monitoring system. The recording setup is attached to the foot-board of the bed, facing the patient (Fig. 2b). Fig. 2.3 shows an example of the captured scene using both color (blurred for patient privacy) and depth (false coloring) cameras.

2.2.3 Recording Paradigms

Movement Task

A verbally-cued motor task was recorded to provide simple, structured behaviors as a precursory analysis to continuous tracking. Patients were given verbal instructions to perform simple single-jointed movements, such as raising their arms. Each movement was repeated at least 10 times, and the full task lasted approximately 10 minutes.

Continuous Behavior

Continuous recording was done during a period of the day in which we expected the most active behaviors. For NY531, this included approximately two hours of daytime activity in the afternoon where the patient was eating, reading, and conversing. Continuous recordings were not obtained for NY540.

2.3 Experimental Methods

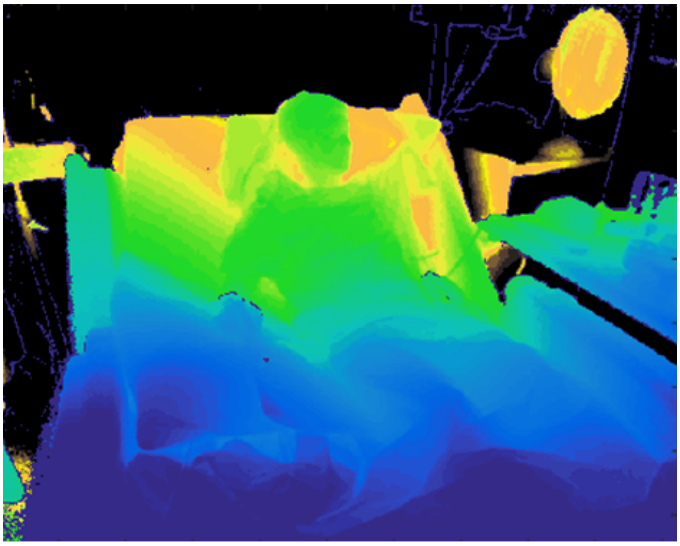
2.3.1 Alignment

Kinect and neural recording hardware were synchronized using an Arduino microcontroller to periodically (120s) send a random number (1:8) of pulses to both systems. The neural system received the signal as a series of square waves, whereas the Kinect microphones recorded the signal as a series of tones. The onset and number of pulses at each sync period was extracted for each system, and the matched sync events were fit to align the data streams. Due to noise in the audio recording, there was uncertainty in the onsets extracted from the Kinect data streams. Recording alignment was therefore limited to 25ms, which is within Kinect video sampling resolution.

2.3.2 Behavior Estimation

Color images were processed to track the position of each patient's hands throughout the recording. An edge-based vision tracking function was used from MATLAB [10]. The tracker used the Kanade-Lucas-Tomasi algorithm [11] to track manually identified regions of interest, such as hand region. Corner point features of the hand regions were detected using a minimum eigenvalue algorithm also from MATLAB. These features were fit into a single cluster, of which the centroid was used to estimate hand position. This method detected gross motor movements of the arms, and on average required eight hours to process every one hour of recorded data.

RGB Frame



Depth Frame

Figure 2.3 **Example of images captured by the Kinect sensor**. For a single scene color, depth (false coloring), and monochromatic infrared (not shown) images were recorded, providing visual and spatial information of the patient. Frame rate for both cameras was 30 fps, though color was halved in low-lighting conditions. The high-definition RGB images have a 1920x1080 resolution, while the depth and infrared images have a 512x424 resolution.

Movement velocity was measured in pixels per second (pps) by dividing the Euclidean distance of hand positions between each frame using the elapsed hardware time between samples. To reduce sampling and tracking noise, the velocities were binned and averaged in 200ms intervals.

Discrete states of movements were assigned using velocity thresholds empirically determined from

the structured paradigm. An additional "non-movement" state was included as a miscellaneous category that likely includes other complex behaviors without arm movements. This was primarily done to reduce noise in extracted behaviors, and to address the sparseness of fast movements.

2.3.3 Neural Conditioning

All ECoG recordings were filtered using finite impulse response notch filters to remove line noise (60, 120, 180, 240 Hz). The resulting time series was high-pass filtered at 1 Hz cutoff frequency to reduce noise. Each channel was then inspected, and electrodes that were either excessively perturbed by noise or showed high variance were discarded. Channels previously identified over ictal populations were also rejected. Finally, electromagnetic noise common to the remaining channels was removed by common average referencing (CAR).

2.3.4 Neural Feature Extraction

Spectral features from the high–γ (70–110 Hz) band of the ECoG signal were extracted by band-passing the conditioned neural data from each channel. The Hilbert envelope was then calculated by taking the amplitude of the analytic signal [12]. The resulting high–γ envelope was then binned every 200ms, and the average signal amplitude was calculated for each bin.

2.3.5 Statistical Test

Neural features were aligned to the estimated behavior. Features were then grouped per channel by movement states, and the Wilcoxon rank-sum test [13] was used to find electrodes with distinguishable neural activity between paired states. Two-sided p-values, p_i, were calculated for each electrode to test the null hypothesis that states were drawn from the same distribution. Given that $m = 22$ channels were tested against this hypothesis, the desired significance, $\alpha = 0.05$, was adjusted using the Bonferroni correction as follows:

$$p_i \leq \frac{\alpha}{m} \tag{2.1}$$

The final value represented the probability that both movement states were drawn from the same distribution. Electrodes were considered significant for distinguishing paired states if $p \leq 0.0023$.

2.4 Results

2.4.1 Tracking Natural Movement

We demonstrate the recording system's capabilities in capturing continuous, naturalistic behaviors of the patient in the clinic. We further show that behavior labels can be semi-automatically extracted from these recordings using computer vision techniques. Fig. 2.4 shows a comparison between the hand speeds estimated over a) the instructed task (8 min), and b) throughout the afternoon (2.25 hours) for NY531. Similar results from the instructed task were found with NY540. Currently, an order of magnitude ($\sim$8x) more "human time" is required to fully complete tracking. The time is primarily spent on re-identifying images perturbed by motion blurs and other recording noise.

2.4.2 Mapping Movements to Electrode Coverage

We used the rank-sum test on each electrode to calculate the probability that samples of high–γ features between paired movement states were taken from the same population. This null hypothesis was rejected for electrodes that satisfied $p \leq 0.0023$, which suggests that the median high–γ power is different between non-movement and right hand movement states. *P*-values were used to estimate the spatial response of cortex to the movement states.

As shown in Fig. 2.5, we focused on the statistical difference between all non-movements and right hand movements taken from the continuous behavior of NY531. The grid corresponds to the red 8x8 array in Fig. 2a. The highlighted electrode is expected to be over motor strip, and its corresponding histogram comparison is shown in Fig. 2.6. As ECoG coverage for the patients was entirely on one cortical hemisphere, we observed statistical significance ($p \leq 0.0023$) in electrodes over pre/post-central gyrus between high–γ magnitudes of total contralateral movement and non-movement states. These areas of cortex correspond to motor planning, action, and sensory regions identified in previous works [14].

2.5 Discussion & Conclusion

We present a minimally intrusive recording system to augment patient neural monitoring. This is a proof of concept that motor behaviors can be estimated alongside hours of continuous neural recording with our system, and can be used to generate functional maps of the brain. At this time, the behaviors and features extracted from these data sets remain noisy. This in part is due to the selection of computer vision algorithms and coarse definitions of natural behavior. We have yet to fully leverage the complexity of movements

recorded in RGB-D video. Our next steps include analyzing more specific subsets of kinematic movements, leverage of the additional depth data stream to reduce motion artifacts, and developing functional maps from

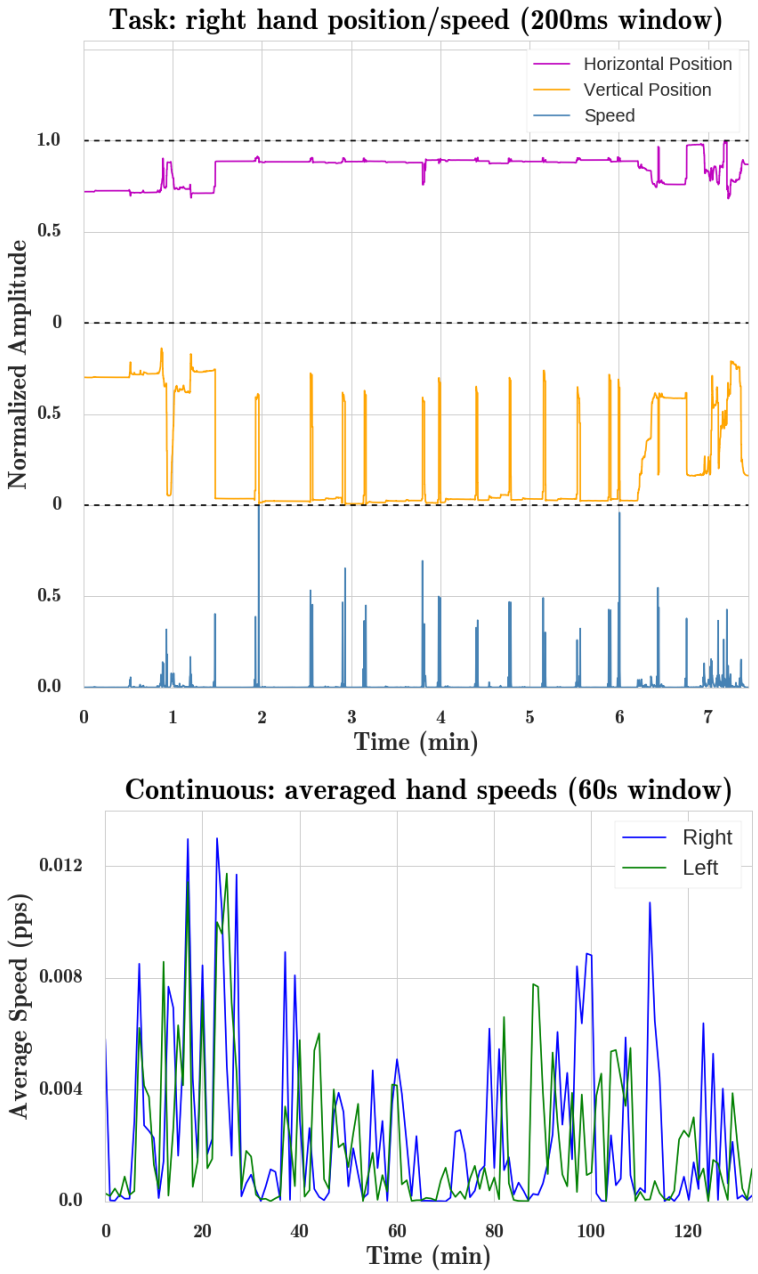


Figure 2.4 **Movement speeds for the patient's (NY531) hands were estimated from both movement task (top) and continuous behavior (bottom)**. Speed was calculated in pixels per second (pps). In the instructed task, the patient was given verbal instruction to perform simple, single-jointed movements, including a stiff forward arm raise. The periodic peaks in movement speed reflect the nature of the task. Similar results were found for subject NY540. For the second plot, a 2.25 hour period in the afternoon was selected for analysis because the patient was awake and active. For visualization, the average speeds in the series were binned every minute.

more stringent statistical tests.

Acknowledgment

This work is supported by the UCSD ECE Department Medical Devices & Systems Initiative, the UCSD Centers for Human Brain Activity Mapping (CHBAM) and Brain Activity Mapping (CBAM), UCSD Frontiers of Innovation Scholars Program, and the Qualcomm Institute Calit2 Strategic Research

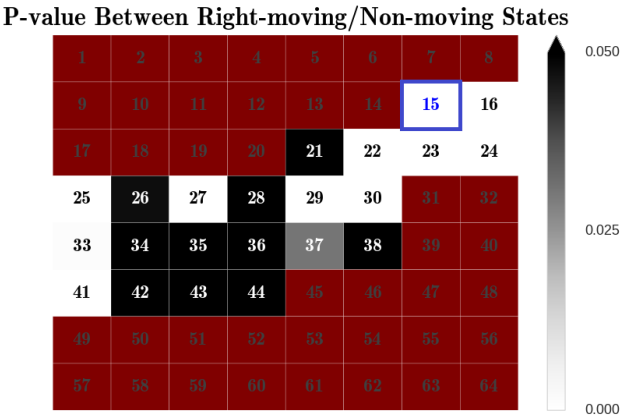


Figure 2.5 **Two-sided p-value for each electrode on the square grid**, equivalent to the red 8x8 array in Fig. 2a. The electrode numbers are displayed over each corresponding cell, and rejected electrodes are marked in maroon. Values were obtained using a rank-sum statistic to compare high–γ (70-110 Hz) amplitude between all contralateral hand movement and non-movement states. The map color scale was capped at $p \leq 0.05$ to emphasize the large difference between electrodes with significant p-values ($p = 2.80\text{e}{-4} \pm 4\text{e}{-7}$; mean $\pm$ st. dev.) and electrodes with non-significant p-values ($p = 0.17 \pm 0.05$).

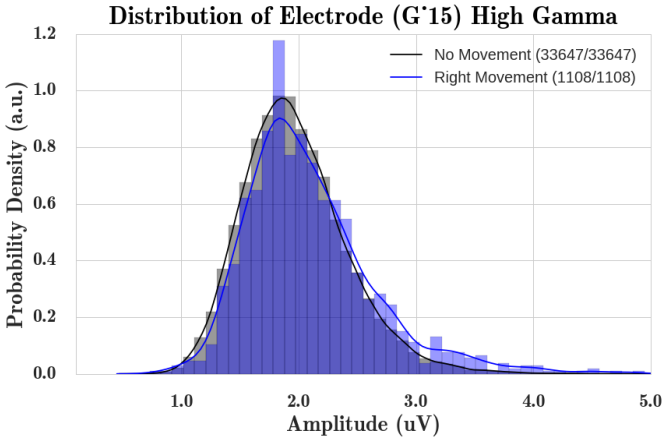


Figure 2.6 **Distribution of high–γ amplitude for two hand movement states (right hand movement, all non-movement)**. Data was taken from the continuous paradigm of NY531, and the highlighted electrode is expected to be over motor strip.

Opportunities (CSRO) program. We would also like to thank Kenny Chen, Preet Minas, and Hugh Wang for their contributions to data collection and analysis.

This chapter is a reprint of the material as it appears in Gabriel, Paolo, Werner K. Doyle, Orrin Devinsky, Daniel Friedman, Thomas Thesen, and Vikash Gilja. "Neural correlates to automatic behavior estimations from rgb-d video in epilepsy unit." In *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference*, pp. 3402-3405. IEEE, 2016.

References

[1] Wang W, Collinger J, Degenhart A, Tyler-Kabara E, Schwartz A, Moran D, Weber D, Wodlinger B, Vinjamuri R, Ashmore R, Kelly J, and Boninger M. An Electrocorticographic Brain Interface in an Individual with Tetraplegia. *PLoS ONE*, **8**, 2013.

[2] Bouchard K, Mesgarani N, Johnson K, and Chang E. Functional organization of human sensorimotor cortex for speech articulation. *Nature*, **495**(7441):327–332, March 2013.

[3] Mesgarani N and Chang E. Selective cortical representation of attended speaker in multi-talker speech perception. *Nature*, **485**(7397):233–6, May 2012.

[4] Dastjerdi M, Ozker M, Foster B, Rangarajan V, and Parvizi J. Numerical processing in the human parietal cortex during experimental and natural conditions. *Nature Communications*, **4**:2528, January 2013.

[5] Jeremy Hill N, Gupta D, Brunner P, Gunduz A, Adamo M, Ritaccio A, and Schalk G. Recording human electrocorticographic (ecog) signals for neuroscientific research and real-time functional cortical mapping. *Journal of Visualized Experiments : JoVE*, (64):3993, 2012.

[6] Felsen G and Yang Dan. A natural approach to studying vision. *Nature Neurosci*, **8**(12):1643–1646, 12 2005.

[7] Jackson A, Mavoori J, and Fetz E. Correlations between the same motor cortex cells and arm muscles during a trained task, free behavior, and natural sleep in the macaque monkey. *Journal of Neurophysiology*, **97**(1):360–374, 2007.

[8] Nancy X. R. Wang, Jared D. Olson, Jeffrey G. Ojemann, Rajesh P.N. Rao, and Bingni W. Brunton. Unsupervised decoding of long-term, naturalistic human neural recordings with automated video and audio annotations, 2015.

[9] Quinn B, Carlson C, Doyle W, Cash S, Devinsky O, Spence C, Halgren E, and Thesen T. Intracranial cortical responses during visual-tactile integration in humans. *The Journal of Neuroscience : the official journal of the Society for Neuroscience*, **34**(1):171–181, January 2014.

[10] Z Kalal, K Mikolajczyk, and J Matas. Forward-backward error: Automatic detection of tracking failures. In *Pattern Recognition (ICPR), 2010 20th International Conference on*, pages 2756–2759, Aug 2010.

[11] B Lucas and T Kanade. An iterative image registration technique with an application to stereo vision. In *Proceedings of the 7th International Joint Conference on Artificial Intelligence - Volume 2*, IJCAI'81, pages 674–679, San Francisco, CA, USA, 1981. Morgan Kaufmann Publishers Inc.

[12] W Freeman. Hilbert transform for brain waves. *Scholarpedia*, 2(1):1338, 2007.

[13] Frank W. Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6):80–83, 1945.

[14] W Li, N Andreasen, P Nopoulos, and V Magnotta. Automated parcellation of the brain surface generated from magnetic resonance images. *Frontiers in Neuroinformatics*, 7(23), 2013.

Chapter 3

Coarse Behavioral Context Decoding

Abstract

Objective. Current brain-computer interface (BCI) studies demonstrate the potential to decode neural signals obtained from structured and trial-based tasks to drive actuators with high performance within the context of these tasks. Ideally, to maximize utility, such systems will be applied to a wide range of behavioral settings or contexts. Thus, we explore the potential to augment such systems with the ability to decode abstract behavioral contextual states from neural activity. *Approach.* To demonstrate the feasibility of such context decoding, we used electrocorticography (ECoG) and stereo-electroencephalography (sEEG) data recorded from the cortical surface and deeper brain structures, respectively, continuously across multiple days from three subjects. During this time, the subjects were engaged in a range of naturalistic behaviors in a hospital environment. Behavioral contexts were labeled manually from video and audio recordings; four states were considered: *engaging in dialogue*, *rest*, *using electronics*, and *watching television*. We decode these behaviors using a factor analysis and support vector machine (SVM) approach. *Main Results.* We demonstrate that these general behaviors can be decoded with high accuracies of 73% for a 4-class classifier for one subject and 71% and 62% for a 3-class classifier for two subjects. *Significance.* To our knowledge, this is the first demonstration of the potential to disambiguate abstract naturalistic behavioral contexts from neural activity recorded throughout the day from implanted electrodes. This work motivates further study of context decoding for BCI applications using continuously recorded naturalistic activity in the clinical setting.

3.1 Introduction

Brain-computer interface (BCI) research has experienced an impressive growth since the first demonstration of direct control of a robotic manipulator by employing electrical activity generated by ensembles of cortical neurons [1]. Now, BCIs designed for experimental and clinical studies can translate raw neuronal signals into control signals for various applications including, but not limited to, motor prostheses [2–7], speech decoding [8, 9], and functional clinical mapping [10]. For many of these studies, performance is optimized by constraining the particular application to be highly structured, so as to reduce behavioral complexity and variability due to exogenous factors. These structured task-based studies, particularly those with human subjects, use limited datasets, typically on the order of minutes to tens of minutes. Limited dataset size often constrains the complexity and generalizability of BCI decoding strategies explored. In an effort to vastly augment data collection, preliminary work has been conducted to study neuronal signals in a more naturalistic, unconstrained manner by leveraging hours of recordings from monkey [11] and patients taken from the epilepsy monitoring unit (EMU) [12, 13].

In that regard, we labeled unstructured behavior from patients in the EMU. As a complement to trial-based experiments, these data are comprised of hand labeled recordings from multiple hours of video recorded in the EMU during which the subjects are engaged in naturalistic activities and are not instructed to engage in specific tasks. These labels are paired with simultaneously recorded electrocorticography (ECoG) or stereo-electroencephalography (sEEG) signals. ECoG captures electrical activity directly across a significant portion of the cortical surface, while sEEG utilizes depth electrodes that record from deeper brain structures as well as the cortical surface. We utilized these datasets to examine the potential to decode abstract behavioral contextual states from recorded neural activity. We classified different broad behavioral descriptions, which are *engaging in dialogue*, *watching TV*, *using electronics*, or *being at rest*, using a dataset of at least 9 hours per subject of neural data with at least 40 minutes of data for each behavioral label. For future BCI applications, it is critical to have reliable prosthetic performance across different behavioral contexts [11, 14]. We show in this paper that these complex and abstract behavioral contexts are separable. In practical terms, these separable contexts can be used to drive a context switch (Figure 3.1) in future, more generalizable and scalable BCIs. These future BCIs must work well across varied conditions automatically [11, 14]. A switch in context may change the dynamics of the observed neural signals and the intended range of control signals; ideally, a BCI will recognize and respond to those changes [11]. For example, a BCI that can detect behavioral context changes could engage context-specific and task-specific decoding algorithms, as shown in Figure 3.1. Such a BCI would, for example, be able to interpret current

behavioral contexts to efficiently use resources, going into a power saving mode when the user is at rest, or to prevent generating unwanted actions, such as moving a limb when the subject intends to speak [11, 14]. Furthermore, a context decoder could be used in conjunction with a deep brain stimulator (DBS) in a closed-loop system for patients with movement disorders. Rather than continuously supplying the brain with current, which could induce unwanted side effects, such as speech and balance impairment [15, 16], and impact battery life [17], a closed loop DBS with a context switch implemented could be an adaptable system that would choose the most optimal stimulation parameters according to context. For example, after patient programming in the clinic, two sets of parameters could be determined, one that would best suppress hand and arm tremors, while another would suppress less of the tremor, but would cause less speech impairment. In a fully integrated DBS and context decoder system, the system would choose the latter set of stimulation parameters when the patient is determined to be engaging in dialogue. Such future potential BCI and DBS designs motivate investigating the separability of behavioral contexts within the recorded neural activity space. The purpose of this study is to provide a proof of concept that these naturalistic behavioral contexts can be decoded from recorded neural activity.

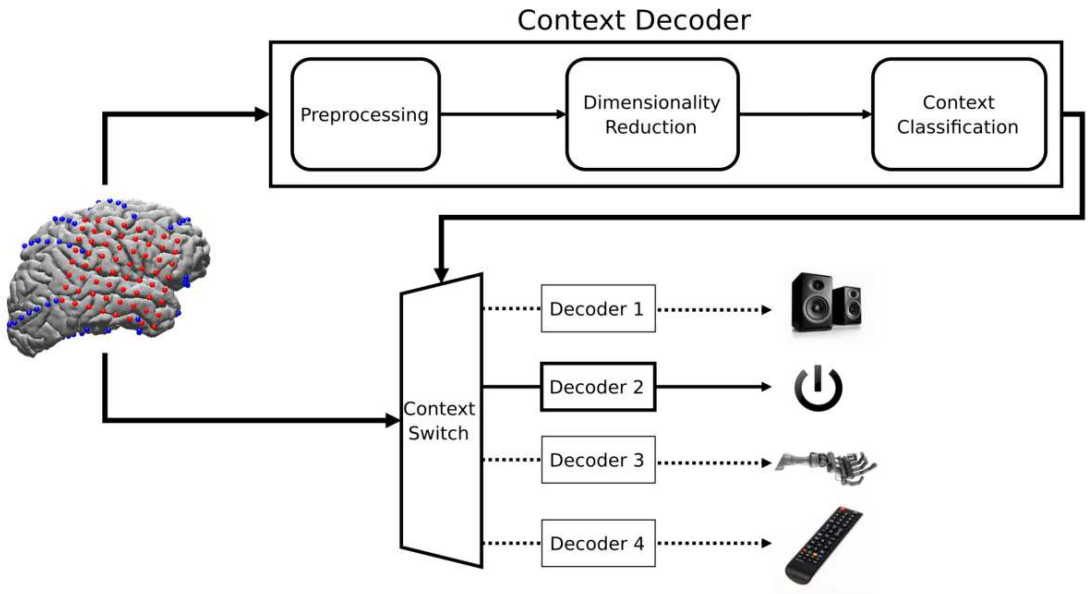


Figure 3.1 **Brain-computer interface (BCI) application for a context decoder.** Neural signals recorded from ECoG and sEEG are processed to drive both a context decoder and a context-dependent BCI application. Current BCIs extract specific features and apply unique decoding strategies on recorded neural activity to achieve high performance in specific tasks. To generalize the capabilities of dedicated BCI applications, we evaluate the possibility of a context decoding layer whose output would be used as a switch between different decoders. A BCI that is able to operate on natural behavior should be able to distinguish between different contexts and select the appropriate task-specific decoder.

3.2 Methods

3.2.1 Description of Data

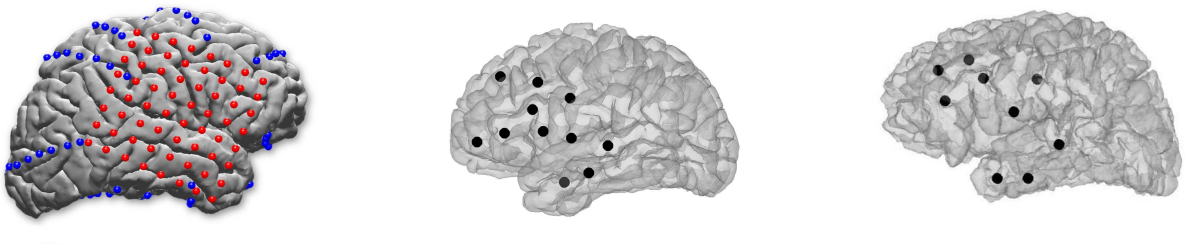
Subjects

Three subjects with intractable epilepsy with hospital ID codes corresponding to NY394, RCH1, and RCH3 participated in this study. For simplicity, they will be respectively referred as Subjects 1, 2 and 3 for the remainder of this work. All subjects underwent invasive monitoring to localize epileptogenic zones before surgical resection. During their stays at New York University (NYU) Langone Comprehensive Epilepsy (Subject 1) and Rady Children's Hospital, San Diego (RCH) Pediatric Epilepsy centers (Subjects 2 and 3), the subjects gave their informed consent and were enrolled in the study. The studies were approved by the Internal Review Board of both institutions.

A Microsoft Kinect v2 was used to record both audio and video for each subject's stay using multiple modalities such as RGB, IR, and depth [12]. However, only audio and RGB video was used for this study. The audio channel was used to synchronize the video and neural streams .

Recording

Subject 1 was implanted in the subdural space with platinum electrode arrays (Ad-Tech Medical Instrument Corporation, Oak Creek, WI) using a combination of linear arrays consisting of 4 to 10 electrode contracts, and a single 8x8 contact grid. Over 100 clinical subdural electrode contacts were embedded in SILASTIC sheets (2.3mm exposed diameter, 10 mm center-to-center spacing) [18] were implanted directly on the right cortical surface across multiple brain regions as shown in Figure 3.2a. The placement of electrodes was based on clinical considerations for the identification of seizure foci. EEG activity was recorded in the frequency range from 0.1 to 230 Hz using Nicolet clinical amplifiers, digitized and sampled at 512 Hz, and referenced to a 2 contact electrode array with electrodes screwed into the skull at a cm distance from the craniotomy edge under the scalp. For this study, we focused on recordings from the 8x8 grid given its coverage of frontal and temporal lobes (Figure 3.2a). These cortical regions of interest have been shown to be engaged in motor [19] and language [20] related behaviors, respectively. Subjects 2 and 3 were implanted with stereo-electroencephalography (sEEG) electrodes via the ROSA robotic surgical implantation (Medtech Surgical Inc., USA) in various orientations to allow for intracranial recording from lateral, intermediate, and/or deep cortical and subcortical structures in a 3-D arrangement [21]. sEEG electrode contacts had a 0.8 mm diameter, were 2mm long, and spaced 1-1.5 mm apart. Electrodes were



(a) *Subject 1* (b) *Subject 2* (c) *Subject 3*

Figure 3.2 **ECoG grid and sEEG shanks locations.** Subject 1 has an ECoG grid with broad right hemisphere coverage. Subject 2 has 11 sEEG shanks placed on the left frontal and temporal lobes. Subject 3 has 9 sEEG shanks placed on the left frontal and temporal lobes.

sampled at 2000 Hz using Xltek 128Fs clinical amplifiers. For Subject 2, the sEEG implants were placed on the left frontal and temporal lobes (Figure 3.2b) while for Subject 3 the implants were placed on both the left and right frontal and temporal lobes. Only electrodes implanted in the left hemisphere for Subject 3 were used (Figure 3.2c).

Table 3.1 Description of data

Subject ID	**Hospital ID**	**Age**	**Sex**	**Duration (hrs)**	**Implant**	**Location**	**Electrodes used**
1	NYU394	30	Male	36	ECoG	Right Hemisphere	51
2	RCH1	18	Female	9	sEEG	Left Frontal and Temporal	71
3	RCH3	19	Male	10	sEEG	Bilateral Frontal and Temporal	68

Labeling

For each subject, we analyzed simultaneously recorded neural signals and clinical video/audio recording began 1-3 days (post-implant) of the subjects' stay in the hospital. Hours of continuous recordings were divided into 5-minute blocks and manually labeled according to the predominant behavior observed during each video segment. A behavioral rubric was designed for this study, which categorized waking periods according to whether or not the subject was active and, if so, who or what they were interacting with. For this analysis, a subset of behavior labels were selected to represent 1 inactive ('Rest') and 3 active

Table 3.2 Labeling criteria for behavior

Label	**Description** (subject is...)	**Mins S1**	**Mins S2**	**Mins S3**
(D) Dialogue	engaging in dialogue (including when the subject is talking on the phone)	275	75	50
(EL) Electronics	using an electronics device (phone, remote, tablet, video games, etc).	40	130	210
(TV) Television	watching television (does not include using remote or watching video on a tablet or phone)	90	N/A	N/A
(R) Rest	awake but resting, either with eyes open or closed, and either with or without visitors in the room	470	165	140

Table 3.3 Labeling criteria for time of day

Label	**Time**	**Mins S1**	**Mins S2**	**Mins S3**
(E1) Epoch 1	2:00-12:59	340	N/A	N/A
(E2) Epoch 2	13:00-16:59	150	125	185
(E3) Epoch 3	17:00-20:59	160	155	165
(E4) Epoch 4	21:00-1:59	225	90	50

('Dialogue', 'Electronics', 'Television') behaviors. Electronics use included using the phone, remote, tablet or playing video games. Additionally, when a subject is talking to someone on the phone, that behavior is labeled as 'Dialogue', not 'Electronics'. Also, the 'Television' label does not include the times when the subject is using the remote. We used 3 active states and 1 'Rest' state in our context decoding scheme for Subject 1, but only 2 active states ('Dialogue', 'Electronics') and 1 rest state for Subject 2 and 3. This is due to the lack of the 'Television' context when labeling Subject 2 and 3's behaviors throughout the day. We also observed that different contexts tend to happen at specific times. For example, when considering time after 22:00, there is a very high chance that the subject is at 'Rest'. Therefore, we applied a second set of labels to relate behaviors to time of day. Note that the non-uniform segmentation of time is due to the nature of the dataset in which interesting contexts are more prevalent at different times throughout the day. Tables 3.2 and 3.3 show a summary of behavior and time labels. Subject 1's behavioral labels were verified

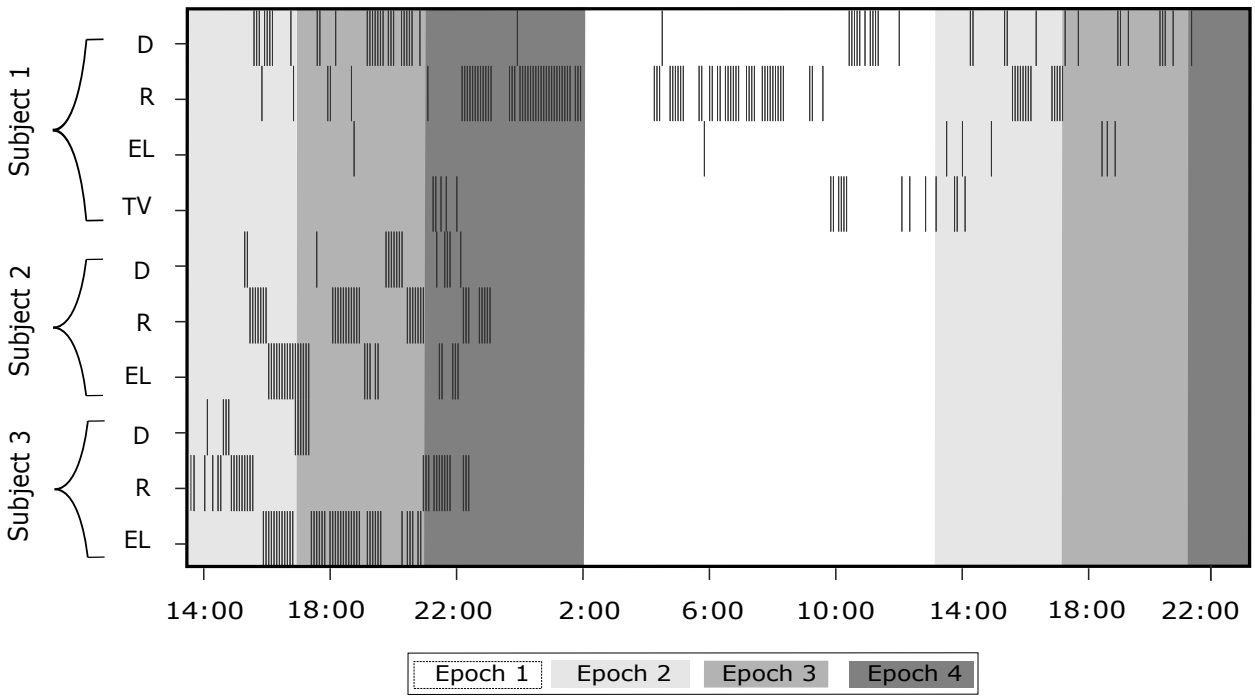


Figure 3.3 **Raster plot showing the locations of each context in time.** Each behavior label represents 5-minutes of manually annotated video. Epochs were segmented around periods where behavior labels were more prevalent throughout the day.

by a second author and Subject 2 and 3's behavioral labels were not verified by a second author.

3.2.2 Neural Signal Conditioning

Prior to analysis, ECoG channels that were over seizure foci were down-selected manually by a clinician based upon visual inspection for Subject 1, while the sEEG channels for Subjects 2 and 3 did not undergo manual down-selection. Furthermore, we have rejected sEEG channels that were outside the cortical surface. Additional ECoG and sEEG channels were rejected if their mean of the squared signal value exceeded a threshold of 3 or 1.5 standard deviations (for ECoG and sEEG respectively) above the mean squared signal value across all channels, indicative of potential noise corruption. sEEG channels were then down-sampled from 2000 Hz to 500 Hz for faster processing. The remaining channels were high-pass filtered with a cutoff frequency of 1 Hz to remove any underlying DC signal and then notch filtered with center frequencies at 60, 120, 180, and 240 Hz to remove line noise and its harmonics. Finally, electrode channels were re-referenced, either to a common average of all channels to remove shared noise (ECoG) or in a pair-wise manner with respect to neighboring electrodes (sEEG).

Spectral features from the 8–25 Hz and 70–110 Hz band of the ECoG signal were extracted by

band-passing each channel. We have only utilized the high frequency band for the sEEG signal due the bipolar referencing scheme that was implemented which is likely to suppress spatially broad low frequency activity. Previous BCI studies motivate the selection of these frequency bands, as they demonstrate power modulation with respect to sensory-motor behavior [22–24]. The signal envelope was then calculated by taking the amplitude of the analytic signal of the Hilbert Transform. The resulting envelopes were then binned every 3 seconds and the average signal amplitude was calculated as an estimate of the square root of the power in each bin. Each channel was then z-scored across time.

3.2.3 Dimensionality Reduction: Factor Analysis

In the factor analysis model, we assume that there exists a multivariate latent random variable that has a Gaussian distribution with zero mean and identity covariance. The model assumes that each high-dimensional neural sample x is generated by sampling a lower dimensional multivariate Gaussian z. Then z is mapped to a higher k-dimensional space by computing $\mu + \Lambda z$, where μ is the observed mean. Lastly, x is generated by adding a strictly diagonal covariance noise Ψ to $\mu + \Lambda z$. We therefore obtain the following probability distributions:

$$z \sim \mathcal{N}(0, I) \tag{3.1}$$

$$x|z \sim \mathcal{N}(\mu + \Lambda z, \Psi) \tag{3.2}$$

Factor analysis is a dimensionality reduction tool similar to principle component analysis (PCA). However, unlike PCA, factor analysis takes into account sensor noise (in this case ECoG or sEEG electrode noise) that is modeled by the diagonal matrix Ψ and allows the scale of sensor noise to vary across sensors. If sensor noise is non-uniform, the principle components found by PCA are biased towards the axes of the noisiest sensors. Thus, when sensor noise can vary, like in ECoG or sEEG, factor analysis can model this non-uniformity and avoid this biasing issue to more effectively find common structure across sensors. We used the Laurens van der Maaten dimensionality MATLAB toolbox for factor analysis [25] which implements the expectation-maximization based algorithm described in [26] to find Λ and Ψ. The joint distribution of x and z is as follows:

$$\begin{bmatrix} z \\ x \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ \mu \end{bmatrix}, \begin{bmatrix} I & \Lambda^T \\ \Lambda & \Lambda\Lambda^T + \Psi \end{bmatrix} \right) \tag{3.3}$$

We projected the electrode space into a factor space of 30 dimensions by calculating the conditional expectation $E(Z|X)$. Since both x and z are random variables obtained from a Gaussian distribution, the conditional

distribution is also Gaussian and can be calculated analytically based upon the joint distribution above, thus $E(Z|X)$ evaluates to $\Lambda^T(\Lambda\Lambda^T + \Psi)^{-1}x$.

3.2.4 Classification

These projected factors form the input features for both a 4-class support vector machine (SVM) classifier with labels 'Dialogue', 'Electronics', 'TV', and 'Rest' for Subject 1. A 3-class support vector machine with labels 'Dialogue', 'Rest', and 'Electronics,' was used for Subjects 2 and 3. We also use a 4-class SVM in classifying time of day labels, 'Epoch 1,' 'Epoch 2,' 'Epoch 3,' and 'Epoch 4' for Subject 1 and a 3-class SVM using time of day labels 'Epoch 2,' 'Epoch 3,' and 'Epoch 4' for Subject 2 and 3. The multi-class SVM uses a one-vs-one coding design and a linear kernel. SVM classification was implemented using the MATLAB machine learning toolbox. We generated confusion matrices for the 4-class and 3-class classifiers depending on the number of contexts available for each subject. We applied a 7-fold cross validation on the data, where the test and training data were each taken from a continuous time segment rather than randomized points across the entire dataset. This was done to minimize the chance that classification was achieved due to the temporal locality of the testing and training data points. The number of samples for each class in the training set is randomly downsampled to the number of samples in the class with the least number of samples to ensure that the SVM does not skew in favor of predicting the majority class. The test set class samples were also balanced to simplify comparison to chance performance. For example, looking at Subject 1's data in Table 3.3, the label with the least amount of data is 'EL' with 40 minutes. Therefore, each fold for electronics would have around 5.7 minutes worth of neural data, which means that we have around 34.2 minutes per class for training and 5.7 minutes per class for testing in each fold.

3.2.5 Electrode Analysis

To motivate future BCI design in terms of implant location on the brain, we ranked the electrodes for each subject by sorting the single channel context decoding performance for each electrode using the classification scheme described in Section 3.2.4. The ranked channels for each subject were then used to investigate the effect of an incremental addition of the best performing electrodes. The accuracy vs. the number of top channels used were plotted for each subject. Finally, we investigated if neighboring electrodes carry similar information that discriminate between different contexts. This was done by segmenting the electrodes into groups belonging to either 2x2 clusters (for Subject 1 who is implanted with an 8x8 ECoG array, giving us 16 different clusters), or shanks (for Subjects 2 and 3 who were implanted with sEEG

shanks). We then found the decoding performance for each cluster/shank and compared them with each single electrode that belongs to that cluster or shank. We plotted the single channel performance versus the cluster/shank performance for each subject. If neighboring electrodes carry the same information regarding context separability, then we should see single channels perform as well as the clusters or shanks that they belong in.

3.2.6 Modeling Neural Activity Dependence on Time

Having observed that different behavioral contexts tend to happen at specific times of day, a possible confound to the context decoder described is that it is not truly decoding context from neural activity, but instead the decoder could rely upon neural activity variation that is primarily coupled by time of day. Thus, we investigated the influence of *temporal distance* and *time of day*. Figure 3.4 is a simple illustration of three possible graphical models that could describe the relationship between time of day, behavioral context, and neural activity generation and their affects on context classification performance. Model 1 suggests that the ability to decode natural behaviors is completely dependent upon the time of day confound rather than behavioral contexts, as it suggests that neural activity is generated based upon time of day and not context. Model 2 and 3 both suggest that context can be disambiguated from neural activity, without a reliance on time of day based modulation. In these two models neural activity is directly context dependent, with time of day having a modulatory effect on neural activity in Model 3. These models do not make any claims on how the brain actually works, but rather highlight the mechanisms that are driving our context classifier's performance. Two control analyses were conducted to investigate the potential time of day confound with respect to these three models.

Temporal Distance vs Performance

Temporal distance is defined by the time separation between the training and testing data. A main concern when analyzing this type of data is the effect of temporal distance on the context decoder performance. We investigated the possibility that decoder performance would decrease as temporal distance increases. If this phenomenon proves to be true, it would highly suggest that Model 1 in Figure 3.4 is the underlying model explaining the separability of neural data in the context decoder. It would also suggest that time dependent neural activity is being decoded, rather than context dependent neural activity. We ran two analyses to explore this idea. Firstly, we organized our dataset so that each context was sorted by time stamp. We then trained a model using a temporally adjacent 20 minute segment per class and tested on 5 minute segments per class that varied in temporal distance from the training set. For example, the classifier

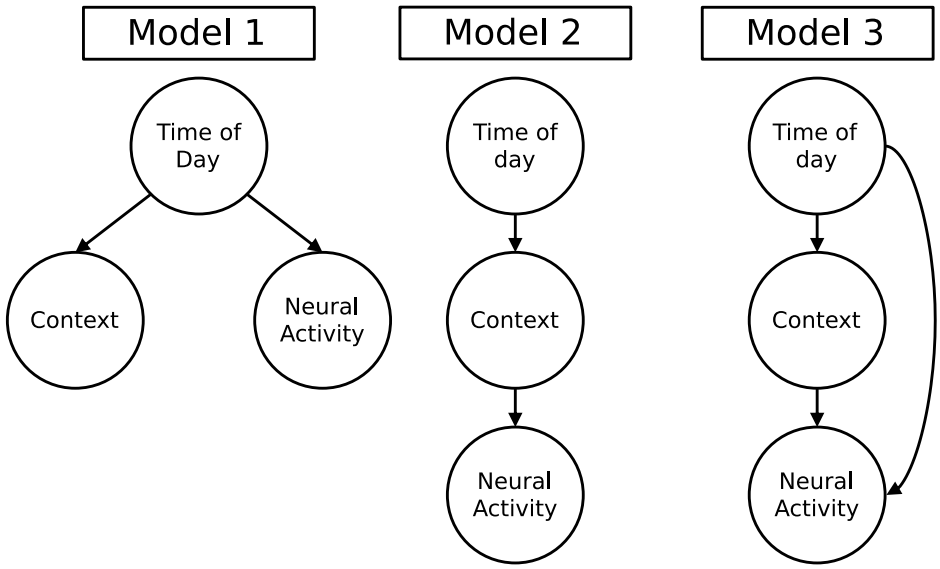


Figure 3.4 **Possible models of time dependencies for behavior context and neural activity that explain context decoding performance.** Model 1 assumes that neural activity and behavioral contexts are independent of each other, but both are dependent on time. Model 2 assumes that the neural activity is dependent only on context, while the context is dependent on the time of day in which it occur. Model 3 assumes that neural activity is dependent on both context and time of day. We did not illustrate these models to provide a generalized explanation of the mechanisms underlying neural activity, but rather to elaborate possible reasons for classification performances when applying a context decoder on a set of neural data.

could be trained from a time segment from 15:00 to 15:20 and tested using a time segment from 16:00, 18:00 and 20:00. We tested the hypothesis that temporal locality is being leveraged by evaluating whether a statistically significant downward trend was found between the temporal distance between the training and test sets and performance. This process was repeated 30 times for different temporal distances. Additionally, to prevent over-fitting due to the small amount of training data compared to context decoding using the entire dataset, we determined the top performing electrodes by sorting the single channel performance using the method described in Section 3.2.4. We ran the temporal analysis using the top 5 electrodes. Linear regression was used to fit a line on the data for each iteration of the temporal analysis to quantify the existence of any statistically significant trend between accuracy and average temporal distance. For this specific analysis, we z-score the test data using the test data's mean and standard deviation to prevent trending effects influencing our results. If the p-value is greater than 0.05, we can accept the null hypothesis that the slope of the line being fitted is zero and reject the hypothesis that there is a linear trend relating temporal distance and classifier performance. Finally, we investigated the effects of dropping the 'Rest' state from this analysis to see if it alleviated any negative trending effects correlating performance and temporal distance. This is due to the fact that the 'Rest' state is not as well defined as other contexts. For example, a subject being

inactive alone or with the presence of people would both be labeled as 'Rest'.

Time of Day and Context Maximum Likelihood Estimation

We further explored the dependence of neural activity on time of day by exploring three separate classification schemes. We implemented a similar classification approach as mentioned in Section 3.2.4: 3 second time windows and 30 latent factors were used with a linear, multi-class SVM classifier, where behavior labels were substituted by the time of day (Table 3.3). This was compared to a time of day maximum likelihood estimator from the context labels. For this estimator, only the context and time of day labels for the 5-minute blocks are used as the features. This was done by calculating the probability of a specific time of day given a context label and choosing the time of day with the highest probability given a specific context. The performance of the maximum likelihood estimator serves as a conservative baseline to determine the amount of information shared between the time of day and context labels. Additionally, the maximum likelihood estimator serves as a biased estimator (favoring the maximum likelihood estimator) since the same dataset is used for training and testing. We compared the performances of the SVM time of day classifier to the time of day maximum likelihood estimator described above to investigate to what extent the time of day affects the context decoder performance. Finally, we constructed another maximum likelihood estimator that predicts context given the time of day using the context and time of day labels for each 5-minute block as features. This estimator is compared to the context decoder described in the Section 3.2.4. For both of the maximum likelihood estimators, we randomly sub-sampled each class label to have the same number of labels for each class for an easier comparison to chance level. The maximum likelihood time of day and context estimators are governed by the following equations (D, C, and T denote decoder, context and time respectively) :

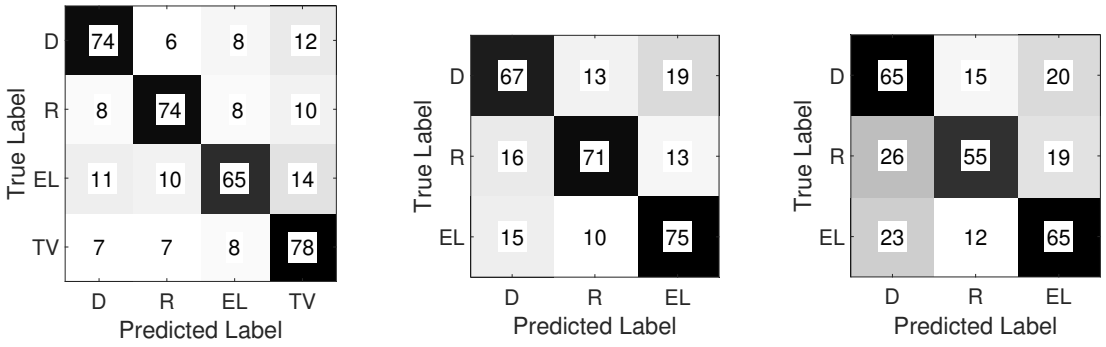
$$D_{context}(j) = \underset{i}{\text{argmax}} P(C_i|T_j) \tag{3.4}$$

$$D_{time}(i) = \underset{j}{\text{argmax}} P(T_j|C_i) \tag{3.5}$$

3.3 Results

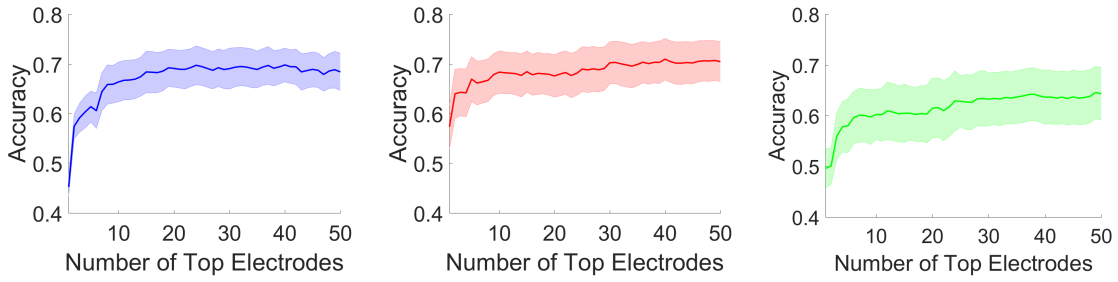
3.3.1 Performance of Classifier

The mean of the confusion matrices for all subjects across folds are shown in Figure 3.5. The linear SVM-based classifier was able to decode different neural states with accuracies significantly above chance performance for all subjects. The classification accuracies when all contexts were used (4-class for Subject



(a) *Subject 1* (b) *Subject 2* (c) *Subject 3*

Figure 3.5 **Confusion matrices for multi-class context decoding using 7-fold cross validation**. Each number in the confusion matrices corresponds to percentages. Rows of the confusion matrices are normalized. Rounding errors will cause some of the rows' sum to deviate slightly from 100%. Both training and test sets are balanced by randomly subsampling the same number of data points in each test fold for each class. For Subject 1, Figure 3.5a, 'Dialogue', 'Electronics', 'TV', and 'Rest' behaviors were classified by a linear, multi-class SVM classifier, using spectral features from the 8-25 Hz and 70-110 Hz bands of ECoG signals collected simultaneously across the brain in 3-second time bins. For Subjects 2 and 3 in Figure 3.5b and Figure 3.5c, only the 70-110 Hz band of the sEEG signal was used. For all three Subjects, behavioral labeling was done in 5 minute blocks. For Subject 3, 'Dialogue', 'Electronics', and 'TV' show a high rate of confusion with 'Rest.' The mean and standard deviation of context decoding accuracy across folds for all subjects can be seen in Table 3.4.



(a) *Subject 1* (b) *Subject 2* (c) *Subject 3*

Figure 3.6 **Accuracy vs Number of Top Electrodes** Each bold line is the mean across 7 folds, and the shaded region is one standard error from the mean across folds. For all 3 subjects, only a small number of top electrodes are needed to achieve adequate context classification performance. The chance level is 25% for Subject 1 and 33% for Subjects 2 and 3.

1 and 3-class for Subjects 2 and 3) were 73%, 71% and 62% for Subjects 1, 2 and 3 respectively. The standard deviation across folds are 9.5%, 10.9%, and 13.5% for Subjects 1, 2 and 3 respectively. We also observe that for Subject 3 in Figure 3.5c, the active states are highly confused with the 'Rest' state.

3.3.2 Electrode Analysis

Looking at Figure 3.6, we observe that the performance of the classifier saturates after a small number of top electrodes are used. Figures 3.7, 3.8 and 3.9 indicate that the top 10 electrodes that contribute most in context decoding occupy multiple regions of the cortical surface (for all subjects) as well as deeper brain structures (for Subjects 2 and 3). Additionally, Figure 3.10 compares the performance when a single electrode is used vs when the entire cluster or shank that electrode is in is used. We observed that for most cases, the 2x2 clusters in the ECoG grid and the sEEG shanks generally outperformed single channels.

3.3.3 Temporal Distance vs Performance

In Figure 3.11, we observe that there is a negative trend relating performance and temporal distance for Subjects 1 and 2, with only Subject 1 having a statistically significant negative trend (p¡0.05). For Subject 3, however, we observe that the fitted slope is almost zero, and becomes slightly positive when we drop the 'Rest' state. Temporal distance has a lesser affect when only the active states were decoded ('Rest' is dropped) for Subject 1 and 2. The fitted slope for Subject 1 when 'Rest' is dropped has a high enough p-value, 0.093, that indicates that the fitted slope is not statistically significant. For all Subjects 1 and 2, the fitted slope is closer to zero and the p-value is greater for when the active contexts are classified compared to when all contexts are classified. We also notice that for certain pairings of training and test sets, the

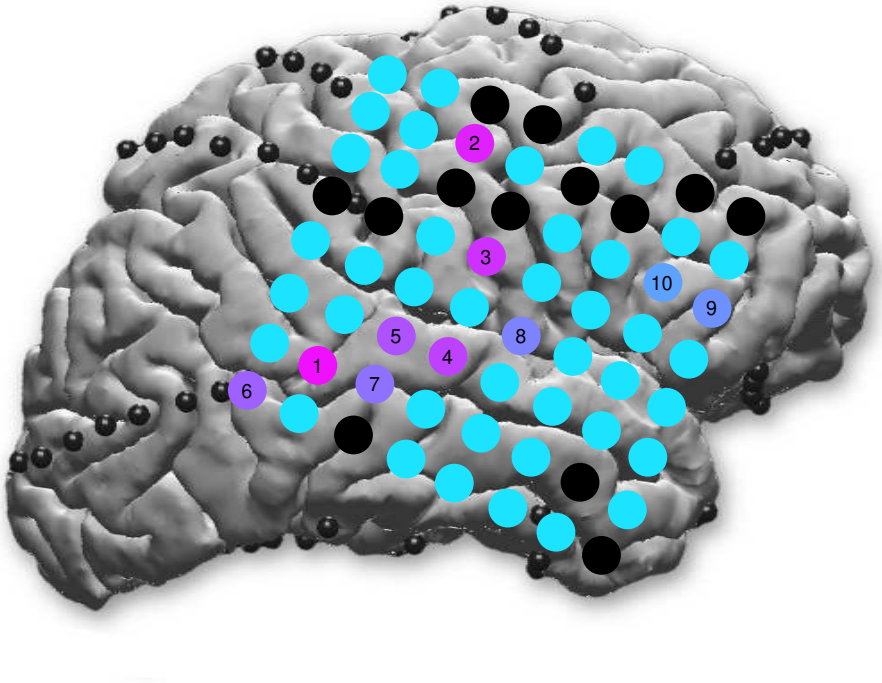
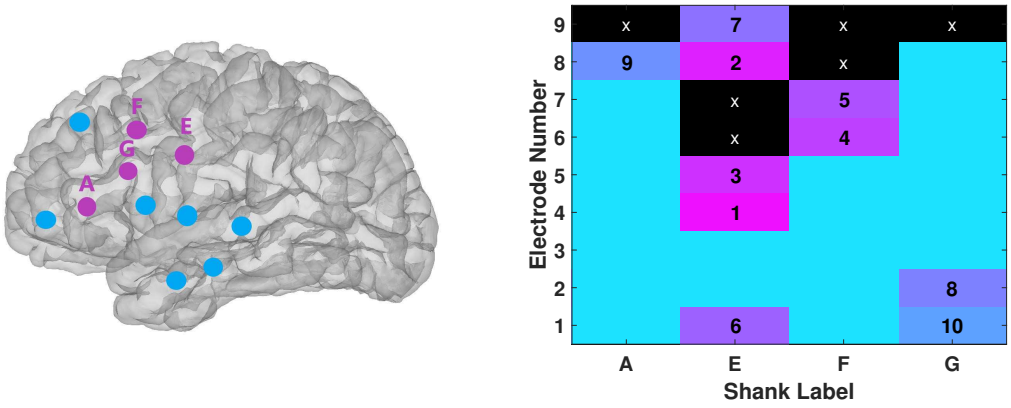


Figure 3.7 **Subject 1**: Locations of the top 10 performing electrodes are numbered. Colors closer to bright pink symbolizes the highest performing electrodes while colors closer to light blue symbolizes the lowest performing electrodes. The electrodes labeled as 'rejected' are colored black.

classifier performance is lower than chance.

Time of Day and Context Maximum Likelihood Estimation

Figure 3.12 shows the confusion matrices of each of the classifiers described in Section 3.2.6 for all subjects. Table 3.4 summarizes the classification accuracies of each type of decoder. The estimator accuracies for the maximum likelihood context estimator given time of day are 38%, 49%, and 51% for Subjects 1, 2 and 3 respectively. In comparison to the context decoder from neural activity accuracies of 73%, 71%, 62% for Subjects 1, 2 and 3 respectively; the context decoder from neural activity outperforms the maximum likelihood context estimator, this suggests that context is separable from neural activity. For the time of day decoders, we observe that the classification accuracies of the maximum likelihood time of day estimator given context are 47%, 49%, 53% and the mean classification accuracies of the time of day given neural activity decoder are 57%, 68%, and 74% with standard deviations across folds of



(a) (b)

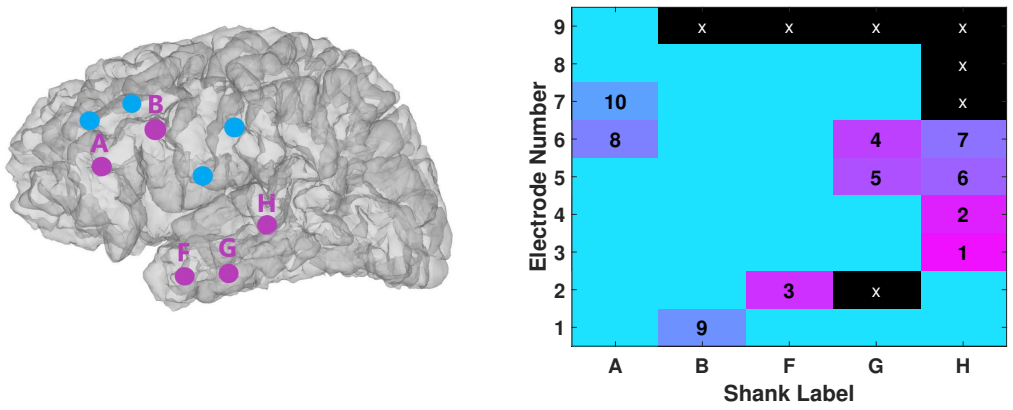
Figure 3.8 *Subject 2*: Figure 3.8a shows the locations of the sEEG electrode shanks that include the top 10 performing electrodes. These shanks are labeled and colored in pink. Figure 3.8b displays the locations of the top 10 performing electrodes across all shanks. Colors closer to bright pink symbolizes the highest performing electrodes while colors closer to light blue symbolizes the lowest performing electrodes. The electrodes labeled as 'rejected' are colored black. Electrode number 9 corresponds to the electrode that is closest to the cortical surface. The electrode contacts are spaced 1-1.5mm apart, are 2 mm long, and have a 0.8 mm diameter.

10.5%, 14.5%, and 7.4% for Subjects 1, 2 and 3 respectively. Subject 1 exhibits the best performance that supports the claim that context-relevant information is being decoded, as the context decoder from neural data outperforms the maximum likelihood context estimator and that the time of day given neural activity decoder does not out perform the maximum likelihood time of day estimator. For Subject 1, the time of day from neural activity decoder performs worse in comparison with other subjects. We theorize that because Subject 1 has the most data, where the distribution of contexts across time is more uniform, as seen in Figure 3.3, the performance of the time of day classifier given neural activity is significantly less than the other two subjects. In Figure 3.3, for Subjects 2 and 3, the distribution of contexts in time is far from uniform, as we notice a significant clustering of the contexts in specific times of day. For example, we observe that for Subject 3, the 'Epoch 2' time label is also exclusively the 'Electronics' context label.

3.4 Discussion

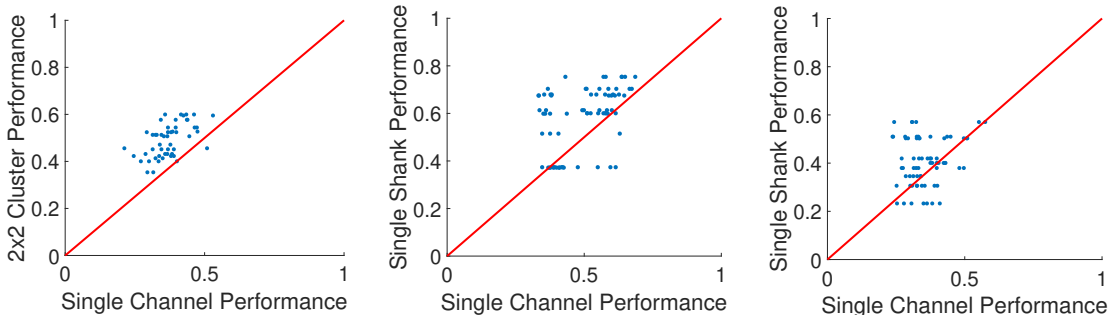
3.4.1 Context Decoding

In this study we demonstrate the potential to decode abstract behavioral contexts from neural activity by achieving classification accuracies significantly above chance for all three subjects as shown in



(a) (b)

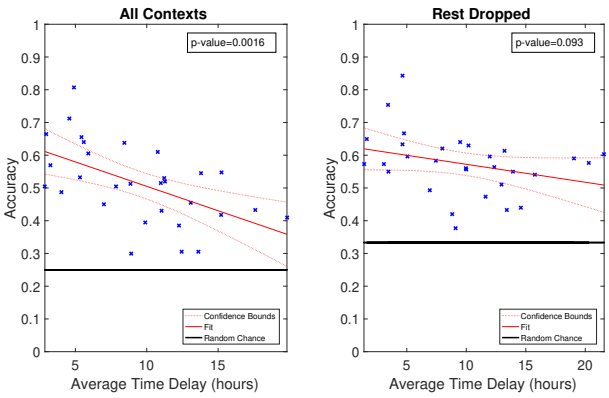
Figure 3.9 **Subject 3**: Figure 3.9a shows the locations of the sEEG electrode shanks that include the top 10 performing electrodes across all shanks. These shanks are labeled and colored in pink. Figure 3.9b displays the locations of the top 10 performing electrodes. Colors closer to bright pink symbolizes the highest performing electrodes while colors closer to light blue symbolizes the lowest performing electrodes. The electrodes labeled as 'rejected' are colored black. Electrode number 9 corresponds to the electrode that is closest to the cortical surface. The electrode contacts are spaced 1-1.5mm apart, are 2 mm long, and have a 0.8 mm diameter.



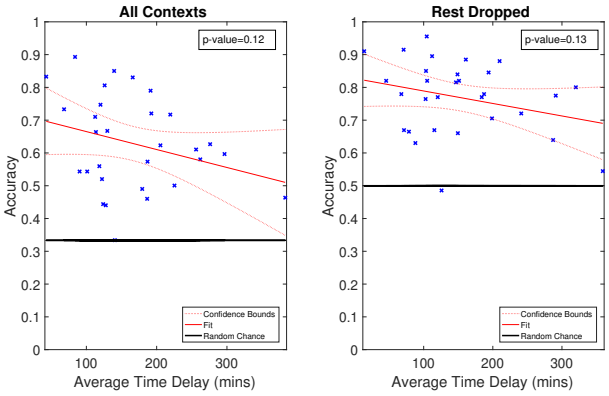
(a) **Subject 1** (b) **Subject 2** (c) **Subject 3**

Figure 3.10 **Shank/Cluster vs Single Channel Performance** For all subjects, the 2x2 ECoG clusters and sEEG shanks out perform single channel performance. This indicates the presence of additive information in neighboring channels, which motivates the use of focal electrode coverage in future BCI design.

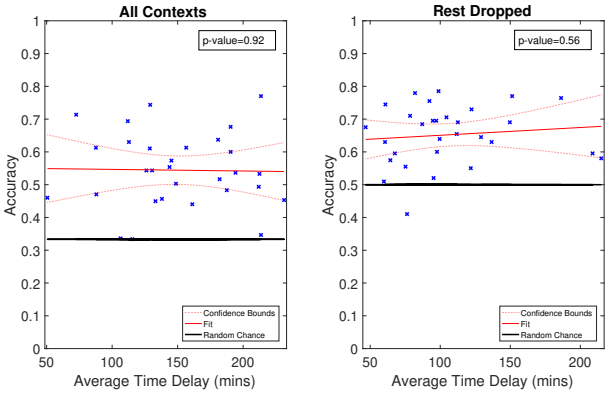
Table 3.4. To our knowledge, this work is the first step in human behavior context decoding from neural signals collected from naturalistic behavior. We discuss in Section 3.1 the practical applications of behavioral context decoding for augmenting BCI and DBS designs. However, the behavioral contexts that we have analyzed in this work is by no means the optimal contexts to implement in a context decoder for a BCI or DBS design scheme. We would like this work to serve as a proof of concept that neural data collected from a naturally behaving subject could facilitate the augmentation of current BCI and DBS designs.



(a) **Subject 1**



(b) **Subject 2**



(c) **Subject 3**

Figure 3.11 **Classifier accuracy vs average temporal distance between training and test sets where k is the number of top electrodes used.** For each subject, we show the possibility of context decoding even when the temporal distance between the training and test set is large. We notice mainly that for Subject 1 in Figure 3.11a, the fitted negative slope statistical significance is greatly alleviated when the 'Rest' state is dropped ($p>0.05$).

Table 3.4 **Performance of different decoders**: Mean accuracies and standard deviations are shown for each decoder for all subjects. Note that for the maximum likelihood estimators, the features were a context or time of day label of a 5-minute segment. Therefore, due to a much reduced sample size as compared to using 3 second bins, k-fold cross validation is not used, and no standard deviations are shown.

Subject	Chance	Context from Neural	MLE-Context	Time from Neural	MLE-Time
1	25%	73% $\pm$ 9.5%	38%	57% $\pm$ 10.5%	47%
2	33%	71% $\pm$ 10.9%	49%	68% $\pm$ 14.5%	49%
3	33%	62% $\pm$ 13.5%	51%	74% $\pm$ 7.4%	53%

3.4.2 Electrode Analysis

The results shown in Section 3.3.2 supports our intuition that to decode complex activities such as *'engaging in dialogue'*, *'using electronics'*, or *'watching TV'*, coverage from multiple regions of the brain is ideal since across three subjects, the top performing electrodes were present in all major regions of the brain in which electrodes were included in the analyses. However, in order to form an adequate statistical analysis of the relationship between different brain regions and contexts, more subjects are needed with broad electrode coverage. We also found that with only a small set of electrodes, we are able to decode context from both the cortical surface and deeper brain regions. In addition to that, we have also found that clusters or shanks generally outperform single electrodes. This indicates that in the space of context decoding, neighboring electrodes in a confined space generate distinctive information that aid in classifying different contexts. Both of these findings suggest that in a practical BCI design, it is beneficial to have broad coverage of different regions of the brain. However, if one has only access to a small portion of the cortex, then a dense electrode coverage would enhance performance.

3.4.3 Temporal Analysis

In Figure 3.11, we show that when 'Rest' is removed, there is less of a negative correlation between temporal distance and performance. This suggests that the 'Rest' state is highly variable and more sensitive to the time of day than other states. Therefore, it is paramount to investigate if there exists sub-states within the 'Rest' state to explore why 'Rest' is highly variable. For all three subjects, we observe many instances where the performance of the classifier is very close to chance or even below chance. This could be due to the fact that we are using a very small dataset for training (20 mins) that is localized in time, which could cause the classifier to overfit for that specific time. However, for many instances, even with limited training

data size per iteration, context decoding is possible even when the temporal distance is large. Also, the precise temporal location of the training and test set might impact this measure.

Our findings in Section 3.2.6 indicate that there is discriminative context-dependent information in the neural signal, which supports Models 2 and 3 rather than Model 1, since the context from neural activity decoder outperforms the maximum likelihood context from time of day estimator for all subjects. However, we also observe that the time of day from neural activity decoder outperforms the maximum likelihood time of day from context estimator for all three subjects as well. Along with the fact that there is a slight negative trend relating decoder accuracy and temporal distance, we conjecture that discriminative information in the neural data is dependent on both time of day and context, which supports Model 3 as the true underlying model that is driving the context decoder's performance. Therefore, additional days of recording are needed to more fully examine the impact of time of day on context decoding. If the performance of the time of day classifier using neural activity was considerably higher than the maximum likelihood time of day from context estimator, then it suggests that the neural activity provides information about time of day beyond the information due to the dependence of context on time of day. In aggregate, the results of these control analyses support Model 3 over Model 1 and Model 2, as described in Figure 3.4. Due to the fact that, first of all, the context decoder from neural activity outperformed the maximum likelihood context estimator from time, and the time of day from neural activity decoder also outperformed the maximum likelihood time of day estimator from context labels. These two observations support the fact that there is both context and time of day information in the neural activity. To fully explore candidate dependencies between neural activity, time of day, and behavioral context, a richer dataset would need to be constructed. Ideally, such a dataset could be sub-sampled to yield datasets in which there is a uniform distribution of different behavioral contexts throughout all time of day labels.

3.5 Conclusion

We have demonstrated the ability to decode coarse and broad behavioral contexts from ECoG and sEEG based neural signals. With a relatively simple machine learning pipeline, we were able to decode behavioral contexts for three subjects with an accuracy significantly above chance. More in-depth analysis and modeling of neural activity is essential to fully understand why these contexts are separable, and to advance our understanding of the underlying physiology. Such analyses could motivate specific electrode coverage paradigms for future BCIs. Furthermore, analysis of context decoders could be used by researchers and clinicians to augment functional mapping of the cortical surface for both clinical and neural engineering

applications with minimal interference to the patient's daily routine. The results obtained from our simple behavioral labeling and decoding scheme motivate further exploration into the methods employed for context decoding. The application of more advanced machine learning methods, such as those that can better leverage temporal and spatial relationships in recorded neural activity, could improve classification performance. This work also motivates the development of richer and more comprehensive datasets of neural activity and naturalistic behavior. Behavioral labeling is a key challenge towards developing such a dataset, as manual annotation is time consuming and subjective. Therefore, there is a need to create an objective and quantitative way to label patient behavior, possibly through video and audio analysis of recorded behavior. Advanced computer vision and speech detection algorithms could be used to generate more descriptive and precise labels, on a timescale of seconds rather than minutes. The development of such datasets would allow for the exploration of richer definitions of contextual state collected across many subjects. In summary, as a proof of concept this work motivates further exploration of behavioral context decoding methods and their potential to advance BCI systems development and clinical neuroscience.

Author Contributions

BCI research is highly collaborative by design, requiring expertise from the fields of clinical neurology, neuroscience, and engineering. To enable this work, the authors contributed in the following ways: Implantation of electrodes was performed via surgery by Drs. Devinsky (NYU) and Gonda (RCH). Clinical monitoring was provided by Drs. Dugan (NYU), Friedman (NYU), Doyle (NYU), Sattar (RCH) and Wang (RCH). Data acquisition was coordinated by Melloni (NYU), Thesen (NYU), Gabriel (NYU,RCH), and mentored by Gilja (NYU, RCH). Behavior labels were annotated by Jiang, Shamie, Alasfour, and mentored by Gilja and Halgren. Analysis and writing were led by Alasfour, assisted by Gabriel, and mentored by Gilja and Halgren.

Acknowledgment

This work has been supported, in part, by the UCSD ECE Department Medical Devices & Systems Initiative, the UCSD Centers for Human Brain Activity Mapping (CHBAM) and Brain Activity Mapping (CBAM), the UCSD Frontiers of Innovation Scholars Program, the Qualcomm Institute Calit2 Strategic Research Opportunities (CSRO) program, the Hellman Fellowship, the Institute of Engineering in Medicine Graduate Student Fellowship, and the Clinical and Translational Research Institute at UC San Diego. We

would like to thank the patients and clinicians who contributed to this study at UC San Diego, Rady Children's Hospital of San Diego, and the Comprehensive Epilepsy Center at NYU Langone Medical Center. We specifically thank Preet Minas, and Hugh Wang for their contributions to data collection and annotation.

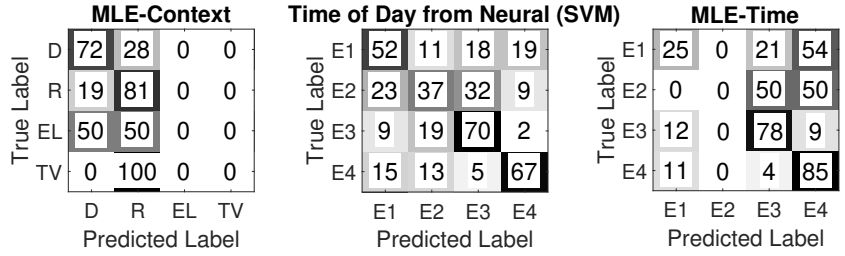
This chapter is a reprint of the material as it appears in Alasfour, Abdulwahab, Paolo Gabriel, Xi Jiang, Isaac Shamie, Lucia Melloni, Thomas Thesen, Patricia Dugan, Daniel Friedman, Werner Doyle, Orin Devinsky, David Gonda, Shifteh Sattar, Sonya Wang, Eric Halgren, and Vikash Gilja. "Coarse behavioral context decoding." *Journal of neural engineering* 16, no. 1 (2019): 016021.

3.6 Supplement

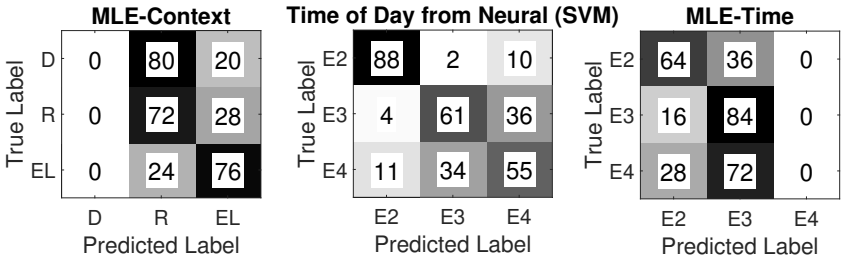
3.6.1 Supplementary Data

Table 3.5 This table shows the amount of data in the training and testing folds for the SVM decoders when applying 7-fold cross-validation (duration in minutes) after randomly down-sampling to have an equal representation of each class in training and testing. See Table 1 for the total amount of data for each class.

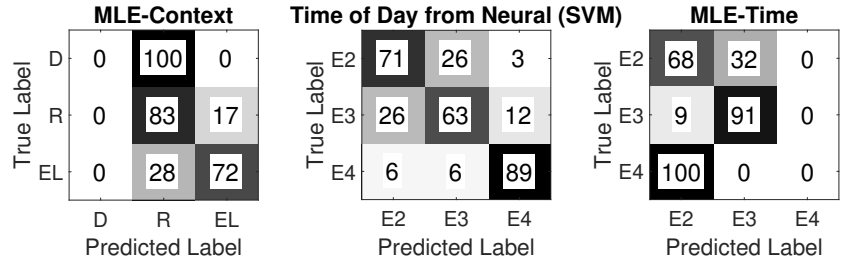
Subject ID	Train/Context	Test/Context	Train/Epoch	Test/Epoch
1	34.3	5.7	128.6	21.4
2	64.3	10.7	77.1	12.9
3	42.9	7.1	42.9	7.1



(a) **Subject 1**



(b) **Subject 2**



(c) **Subject 3**

Figure 3.12 **Confusion matrices for decoding context from time of day, time of day from neural data, and time of day and from context for all three subjects.** Rows of the confusion matrices are normalized. Rounding errors will cause some of the rows' sum to deviate slightly from %100. For all three subjects, (performance summarized in Table 3.4), the context from neural activity out performs the maximum likelihood context from time of day estimator, and the time of day from neural activity decoder also out performs the maximum likelihood time of day from context estimator. However we note that for Subject 1, where context labels are most uniformly spread through time, the time of day from neural activity decoder out performs the maximum likelihood time of day from context estimators the least out of all three subjects. We also note that for maximum likelihood estimators, some labels are completely predicted incorrectly.

References

[1] Fetz E. Operant conditioning of cortical unit activity. *Science*, **163**(3870):955–958, 1969.

[2] Wang W, Collinger J, Degenhart A, Tyler-Kabara E, Schwartz A, Moran D, Weber D, Wodlinger B, Vinjamuri R, Ashmore R, Kelly J, and Boninger M. An Electrocorticographic Brain Interface in an Individual with Tetraplegia. *PLoS ONE*, **8**, 2013.

[3] Carmena J, Lebedev M, Crist R, O'Doherty J, Santucci D, Dimitrov D, Patil P, Henriquez C, and Nicolelis M. Learning to control a brain–machine interface for reaching and grasping by primates. *PLoS Biology*, **1**(2), 10 2003.

[4] Chaplin J, Moxon K, Markowitz R, and Nicolelis M. Real-time control of a robot arm using simultaneously recorded neurons in the motor cortex. *Nature*, **2**:664–670, 1999.

[5] Gilja V, Nuyujukian P, Chestek C, Cunningham J, Krishna S, et al. A high-performance neural prosthesis enabled by control algorithm design. *Nature Neuroscience*, **15**:1752–1758, 2012.

[6] Hochberg L, Serruya M, Friehs G, Mukand J, Saleh M, Caplan A, Branner A, Chen D, Penn R, and Donoghue J. Neuronal ensemble control of prosthetic devices by a human with tetraplegia. *Nature*, **442**:164–171, 2006.

[7] Velliste M, Perel S, Spalding C, Whitford A, and Schwartz A. Cortical control of a prosthetic arm for self-feeding. *Nature*, **458**:1098–1101, 2008.

[8] Bouchard K, Mesgarani N, Johnson K, and Chang E. Functional organization of human sensorimotor cortex for speech articulation. *Nature*, **495**(7441):327–332, March 2013.

[9] Mesgarani N and Chang E. Selective cortical representation of attended speaker in multi-talker speech perception. *Nature*, **485**(7397):233–6, May 2012.

[10] Korostenskaja M, Wilson A, Rose D, Brunner P, Schalk G, Leach J, Mangano F, Fujiwara H, Rozhkov L, Harris E, et al. Real-time functional mapping with electrocorticography in pediatric epilepsy: comparison with fmri and esm findings. *Clinical EEG and Neuroscience*, **45**(3):205–211, 2014.

[11] Santatham G, Linderman M, Gilja V, Afshar A, Ryu S, Meng T, and Shenoy K. Hermesb: A continuous neural recording system for freely behaving primates. *IEEE Transactions on Biomedical Engineering*, **54**(11):2037–2050, 2007.

[12] Gabriel P, Doyle W, Devinsky O, Friedman D, Thesen T, and Gilja V. Neural correlates to automatic behavior estimations from rgb-d video in epilepsy unit. In *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the*, pages 3402–3405. IEEE, 2016.

[13] Wang N, Olson J, Ojemann J, Rao R, and Brunton B. Unsupervised decoding of long-term, naturalistic human neural recordings with automated video and audio annotations. *Frontiers in human neuroscience*, **10**, 2016.

[14] Linderman M, Santhanam G, Kemere C, Gilja V, Shenoy K O'Driscoll S, Meng T, et al. Signal processing challenges for neural prostheses: a review of state-of-the-art systems. *IEEE signal processing magazine*, pages 18–28, 2008.

[15] Baizabal-Carvallo J, Kagnoff M, Jimenez-Shahed J, Fekete R, and Jankovic J. The safety and efficacy of thalamic deep brain stimulation in essential tremor: 10 years and beyond. *Journal of Neurology, Neurosurgery & Psychiatry*, 2013.

[16] Guehl D, Edwards R, Cuny E, Burbaud P, Rougier A, Modolo J, and Beuter A. Statistical determination of the optimal subthalamic nucleus stimulation site in patients with parkinson disease. *Journal of Neurosurgery JNS*, **106**(1):101 – 110, 2007.

[17] Houston B, Thompson M, Ojemann J, Ko A, and Chizeck H. Classifier-based closed-loop deep brain stimulation for essential tremor. In *8th International IEEE EMBS Conference on Neural Engineering*, pages 316–221. IEEE, 2017.

[18] Quinn B, Carlson C, Doyle W, Cash S, Devinsky O, Spence C, Halgren E, and Thesen T. Intracranial cortical responses during visual-tactile integration in humans. *The Journal of Neuroscience : the official journal of the Society for Neuroscience*, **34**(1):171–181, January 2014.

[19] Miller K, Schalk G, Fetz E, Den Nijs M, Ojemann J, and Rao R. Cortical activity during motor execution, motor imagery, and imagery-based online feedback. *Proceedings of the National Academy of Sciences*, **107**(9):4430–4435, 2010.

[20] Pasley B, David S, Mesgarani N, Flinker A, Shamma S, Crone N, Knight R, and Chang E. Reconstructing speech from human auditory cortex. *PLoS Biology*, **10**(1):e1001251, 2012.

[21] Gonzalez-Martinez J, Bulacio J, Thompson S, et al. Robotic epilepsy surgery: technique, results and complications related to robotic assisted seeg. *Neurosurgery*, **78**(2):169–179, 2016.

[22] Arroyo S, Lesser R, Gordon B, Uematsu S, Jackson D, and Webber R. Functional significance of the mu rhythm of human cortex: an electrophysiologic study with subdural electrodes. *Electroencephalography and clinical neurophysiology*, **87**(3):76–87, 1993.

[23] Crone N, Miglioretti D, Gordon B, and Lesser R. Functional mapping of human sensorimotor cortex with electrocorticographic spectral analysis. ii. event-related synchronization in the gamma band. *Brain*, **121**(12):2301–2315, 1998.

[24] Aoki F, Fetz E, Shupe L, Lettich E, and Ojemann G. Increased gamma-range activity in human sensorimotor cortex during performance of visuomotor tasks. *Clinical Neurophysiology*, **110**(3):524–537, 1999.

[25] Van Der Maaten L, Postma E, and Van den Herik J. Dimensionality reduction: a comparative review. *J Mach Learn Res*, **10**:66–71, 2009.

[26] Ghahramani Z, Hinton G, et al. The em algorithm for mixtures of factor analyzers. Technical report, Technical Report CRG-TR-96-1, University of Toronto, 1996.

Chapter 4

Neural Correlates of Unstructured Motor Behaviors

Abstract

Objective. We studied the relationship between uninstructed, unstructured movements and neural activity in three epilepsy patients with intracranial electroencephalographic (iEEG) recordings. *Approach.* We used a custom system to continuously record high definition video precisely time-aligned to clinical iEEG data. From these video recordings, movement periods were annotated via semi-automatic tracking based on dense optical flow. *Main results.* We found that neural signal features (8–32 Hz and 76–100 Hz power) previously identified from task-based experiments are also modulated before and during a variety of movement behaviors. These movement behaviors are coarsely labeled by time period and movement side (e.g. 'Idle' and 'Move', 'Right' and 'Left'); movements within a label can include a wide variety of uninstructed behaviors. A rigorous nested cross-validation framework was used to classify both movement onset and lateralization with statistical significance for all subjects. *Significance.* We demonstrate an evaluation framework to study neural activity related to natural movements not evoked by a task, annotated over hours of video. This work further establishes the feasibility to study neural correlates of unstructured behavior through continuous recording in the epilepsy monitoring unit. The insights gained from such studies may advance our understanding of how the brain naturally controls movement, which may inform the development of more robust and generalizable brain-computer interfaces.

4.1 Introduction

Key advancements in recording hardware, data computation, clinical care, and cognitive science continue to drive new possibilities in how humans and machines can interact directly through thought. Brain-computer interface (BCI) studies show that humans can control computer software and prostheses with brain signals for very specific applications, for example [1–3], and that BCI based tasks can be utilized in the clinic for functional brain mapping [4–6]. The generalization of BCIs across a wider set of behavioral and environmental contexts is largely unstudied. For example, it remains unanswered as to how the human brain controls movements not evoked by tasks. Studies like [7] show neural spiking activity occurring during self-paced motor behavior not induced by tasks that is similar to neural spiking activity induced by a task. But what is unanswered is how unstructured, daily-life motor behaviors compare with specific controlled motor output. In the presented work, we develop a paradigm with methods to enable such studies.

4.1.1 Background

Since the first successful decoding of neural signals to control a robotic manipulator [8], researchers have pushed human brain-computer interfaces BCIs to additional applications, including functional clinical mapping [4–6], speech decoding [9–11], and motor prosthetic control [12–17]. Many experiments were motivated by a greater understanding of neural signal properties [18, 19] and cortical function [20, 21], and were enabled by the availability of greater spatial and temporal resolution in recording neural activity during clinical studies. Particularly, studies like [22] have identified spectral responses to various movement modalities in the high-frequency band (76–100Hz, HFB) and low-frequency band (8–32Hz, LFB) of cortical signals near central sulcus.

Much of our knowledge about BCIs is informed by data collected through carefully controlled experiments in laboratory or clinical environments. Constraining BCI experiments with highly structured paradigms and settings allows researchers to collect a high number of consistent trials for a given behavior, while controlling for external confounding factors, when the timing of recording sessions is limited to several hours or less. By reducing behavioral complexity and variability, researchers have identified specific relationships between the brain and functional behaviors like movement [23] and speech [24]. However, very little is currently known about how well such these task-based relationships extend to daily life [25], in part because collecting data outside of the lab is challenging. To further understand natural brain activity, we must study more complex behaviors in more environmentally relevant settings.

One setting in which such studies are conducted is the epilepsy monitoring unit (EMU). EMUs

admit patients with intractable epilepsy, who undergo invasive monitoring to localize epileptogenic zones prior to surgical resection. During invasive monitoring, these patients have electrodes placed directly on (*i.e.*, subdural) or inside (*i.e.*, depth) their brain. These electrodes allow clinicians to simultaneously sample local field potentials summarizing activity in multiple cortical layers and over cortical surfaces, which can aid in seizure focus localization [26–28] and functional mapping [29]. Implanted patients generally stay in the EMU for 3–14 days to capture seizure activity; outside of brief periods of clinical care, these patients move and act on their own volition, providing the opportunity to continuously monitor their uninstructed and unstructured behaviors. Preliminary work has been conducted to study neural activity in relatively more naturalistic, unstructured settings, from human patients admitted to the EMU. These studies show that neural data collected in such settings can inform movement classification using supervised [30, 31] and unsupervised [32] learning methods. In addition, our recent study [33] demonstrates that from such data, behavior-specific contexts can be discriminated from recorded neural activity.

4.1.2 Challenges

One of the greatest challenges to studying neural activity in unstructured settings is properly curating usable data. Typically, subjects in the EMU are surrounded by medical equipment and clinical staff in an ever-shifting environment. Any of these factors may inject unwanted noise or artifact into either the video or neural recording; task-based studies are typically designed to limit impact of these factors. In addition to these challenges, frequency of activity varies from subject to subject, and periods of movement are relatively sparse compared to periods of rest. Identifying these critical periods while avoiding confounding periods requires careful inspection and annotation. On a small scale, this is usually done by hand, but the task grows in difficulty as longer datasets are recorded. Longer datasets are necessary to capture more examples of behavior and to document a wider range of behaviors. While there are methods to automate certain annotations [34, 35], few are sufficiently reliable in noisy environments. By addressing these challenges, however, one may be able to study neural activity in unstructured settings.

4.1.3 Proposed Method

In this work, continuous multi-modal datasets were collected in the EMU from subjects with electrodes implanted in and on their brain. Hours of video of each subject were recorded alongside neural signals from electrodes across many regions of human cortex. Unlike traditional task-based studies that target highly repeatable and instructed behaviors, these data were collected under naturalistic conditions (*e.g.* subject is stretching, eating, playing), thus providing a richer collection of unstructured human behaviors.

Tools were developed to coarsely annotate periods of movement behaviors from video and to parameterize neural activity with respect to those behaviors.

We hypothesize that this continuous monitoring paradigm will provide examples of complex and nuanced movement behaviors that complement task-evoked movements. Studying neural activity with respect to these more natural movements may inform future BCIs, whereas the movement annotation tools developed for this study may augment behavior monitoring in the clinic. To demonstrate the feasibility of such an approach, this work follows two paths: 1) semi-automated, quantitative tracking and trialization of subject movement from RGB video using available computer vision and signal processing techniques; and 2) application of discrete classification techniques on neural signal features previously associated with instructed movements (LFB and HFB signal power). This work serves as a preliminary demonstration that simultaneously recording intracortical activity and unstructured movements can serve as an additional means to assess neural correlates to human behavior.

4.2 Methods

4.2.1 Subjects

A total of three subjects with intractable epilepsy were enrolled following IRB-approved protocols at two institutions, New York University (NYU) Langone Comprehensive Epilepsy and Rady Children's Hospital, San Diego (RCH) Pediatric Epilepsy centers. Specifically, Subject 1 was enrolled at NYU, while Subjects 2 and 3 were enrolled at RCH. During their stay, each subject underwent invasive monitoring to localize epileptogenic zones before surgical resection. After surgical implantation, each subject received standard clinical care in the EMU of their respective institutions. Informed consent was obtained for each subject prior to running this study. The additional monitoring required for this study did not impact the care that subjects received and played no role in their diagnosis or treatment. At any point, subjects were allowed to excuse themselves from the study. A list of subjects, study recording duration, implant type and locations, handedness, age, and gender is provided in Table 4.1.

4.2.2 Data Acquisition

Multi-modal (audio-video, neural) datasets of each subject were recorded using the experimental setup shown in Figure 4.1a. Neural activity was sampled using intracranial encephalography (iEEG), electrodes placed directly on or inside the brain, while subject movements were recorded with a commercial camera. Data modalities and their collection methods are further discussed in this section.

Table 4.1 List of subjects, their respective IDs according to research protocol, total hours of video recorded, neural implant details, dominant hand, age, and sex. Subjects were implanted with electrocorticography (ECoG) and/or stereotactic electroencephalography (sEEG) electrodes to clinically identify seizure onset and eloquent brain regions. Electrodes were implanted in left (L) and/or right (R) hemispheres to target frontal (FL), temporal (TL), and parietal (PL) lobes.

Subject No.	Patient ID	Recording (hrs.)	Implant type	Coverage	Handedness	Age (y.o.)/Sex
1	NY531	2.25	ECoG/sEEG	LHem.	R	48/M
2	RCH1	9	sEEG	L FL/TL	R	18/F
3	RCH3	24	sEEG	L/R FL/TL	R	17/M

Intracranial Neural Signals

Each subject was implanted with platinum electrode arrays configured as some combination of flat strips (8-10 contacts), 2-D grids (8x8 contacts), or depth (6-10 contact) probes to monitor seizure activity in the brain. The strip and grid electrodes are clinically referred to as electrocorticography (ECoG) electrodes [37], while depth electrodes are referred to as stereotactic (or stereotaxic) electroencephalography (sEEG) electrodes [38]. For each patient, over 100 electrode contacts were implanted across multiple brain regions, as demonstrated for Subject 1 in Figure 4.1b. The placement of electrodes was based entirely on clinical considerations for the identification of seizure foci and mapping of eloquent brain areas. After implantation, electrodes were localized co-registering high-resolution preoperative MRI with postoperative computerized tomography (CT) for the purposes of functional mapping [39, 40].

A technical description of both electrode types used for recording is described in Supplement 1.1. In addition, counts of the number of electrodes selected for the study are provided in Table S1.

Subject Videos

To help clinicians visually identify patient-specific seizure semiologies, EMU patients are also monitored via video recording alongside neural signal acquisition. Due to clinical constraints, these videos are generally recorded from various bedside angles and at a compact resolution. Furthermore, most commercial systems coarsely align video to neural signal by start and end time stamps. These restrictions increased the difficulty of tracking the nuances and timing of subject movement, and warranted an alternative solution.

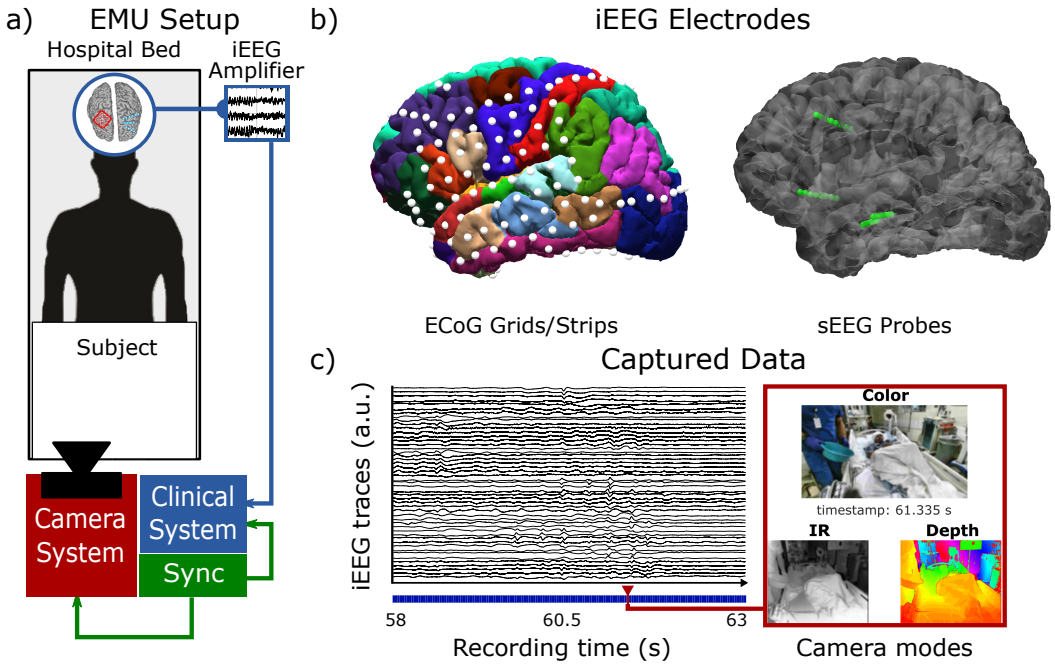


Figure 4.1 **Diagram of experimental setup for recording behavioral video simultaneously with intracortical activity (ECoG, sEEG) from subjects in the epilepsy monitoring unit. a)** The experimental setup places the video recording system at the foot of the hospital bed facing the subject to capture the subject and their immediate surroundings. **b)** A typical subject has over 100 electrodes placed according to clinical need, as shown in the reconstructed coverage of ECoG and sEEG for Subject 1. A combination of subdural grid and strip electrodes (ECoG) covers many regions of the cortical surface, depicted by the Desikan-Killiany parcellation [36], while stereotactic depth probes (sEEG) sample deep and superficial brain structures. **c)** During the study, videos (blue) of the subject moving in an uninstructed and unstructured manner are captured using a Kinect for Windows (v2) sensor. An example of the sensor modalities is framed in red. A subset of 50 neural traces recorded in parallel shown underneath, aligned $\leq 5ms$ of each video frame. Data collected in this manner captures an external and intracortical representation of each subject's movement. A preliminary demonstration using neural features in relation to movement segments marked using only the color video stream is detailed in this work.

A free-standing monitoring unit [30] was designed and built to allow for minimally intrusive video recording of each subject using a Microsoft Kinect v2 sensor. This commercial sensor allowed for simultaneous audio, video, and depth recording, sampled at 16kHz, 30Hz, and 30Hz, respectively. The recording system consisted of the Kinect sensor, a compact computer for data retention, and a custom hardware synchronizer connected to the clinical monitoring system. This synchronizer aligned the video sensor to neural acquisition unit through the audio channel, and is further described in Supplement 1.2. In this work, audio data was used only for alignment. The monitoring unit was placed on the bed end furthest from and facing the subject (Figure 4.1a), in order to capture both the subject and their immediate clinical surroundings.

Recording

Video recording began 1–2 days post-implant when subjects were expected to be more active. The experiment was concluded before the implanted electrodes were surgically removed from the patient, prior to their release from the EMU. A sample of such data is framed in red on Figure 4.1c. Image frames represent the captured scene using color (blurred for patient privacy), infrared (grayscale, also blurred), and depth (false coloring) streams. Subject images are blurred to protect privacy during reporting. No additional processing was applied to video streams during collection. Video recording details, including total number of frames processed per subject, is summarized in Table S2.

4.2.3 Movement Behavior Annotation

Unlike traditional task-based studies that trialize their recording sessions with encoded markers, continuous-monitoring studies require alternative strategies to mark movement periods. Because unstructured movements are not as clearly defined, they require either nuanced details to describe each unique behavior or coarse labels that broadly associate similar types of behaviors. In this work, a dense optical flow algorithm was used to quantify and coarsely label movement from RGB video (Figure 4.2, left column). Movement periods were detected and segmented by thresholding flow magnitude in manually defined regions of interest (ROIs), as described in the subsections below.

Dense Optical Flow

A two-frame motion tracking algorithm was used to quantify the magnitude and direction of movement across RGB video of each subject. The Gunnar Farnebäck's algorithm [41] (OpenCV, *calcOpticalFlowFarneback*) computes the dense optical flow between two sequential frames based on a quadratic polynomial expansion. The resulting optical flow frame (Figure 4.3a) encodes information about the magnitude and direction of movement observed by every pixel, providing more direct detail from the scene than feature-based tracking algorithms [42]. Sequential flow frames provided an estimate for movement across the video, which was used to mark movement periods (Figure 4.3b). Due to artifacts that result from the inconsistent nature of unstructured movement, both magnitude thresholding and clustering methods were applied to the optical flow signal to generate reliable but coarsely labeled movement periods.

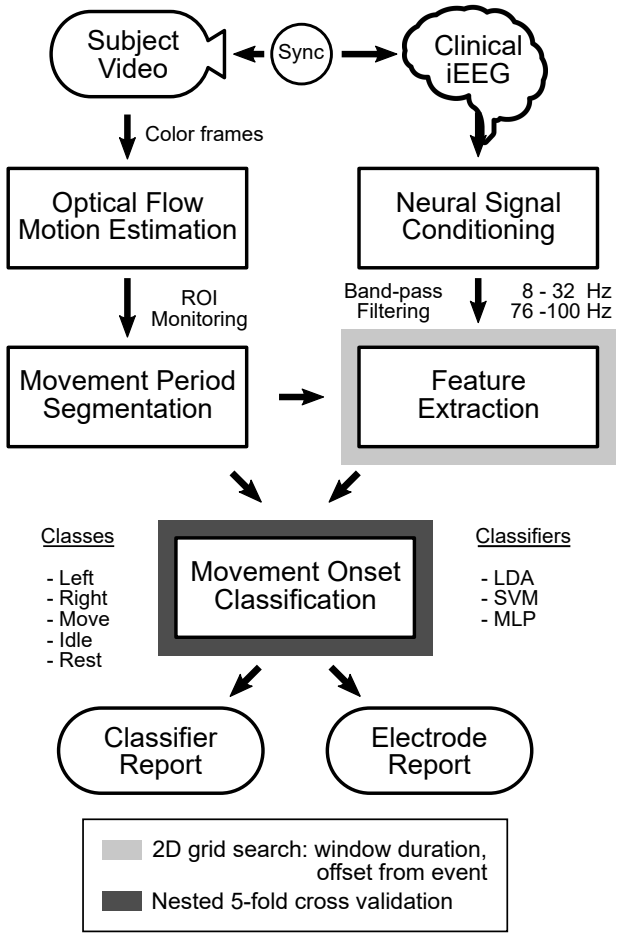


Figure 4.2 **Processing steps of simultaneously recorded color video and subdural activity.** Subject color video and neural recordings were processed separately. (Right) Neural electrode recordings were conditioned and then band-pass filtered into two frequency ranges: 8-32Hz (LFB) and 76-100Hz (HFB). (Left) Estimates of movement activity across video were generated using a region of interest (ROI) template and a dense optical flow algorithm to semi-automatically annotate time segments with of coarse movement labels. Neural samples were binned using windows of fixed duration and offset from movement annotations (light gray), with the final dataset consisting of trials of binned root-mean-square (RMS) power estimates of LFB and HFB amplitude per valid channel. Finally, the separability of each movement class within the neural feature space was assessed using nested 5-fold ($cv = 5$) cross-validation (dark gray) with a set of classifier models and class combinations. The classification results from the best models generated for select class combinations are reported in this work, along with the solo classification accuracy of each electrode, which was calculated during model training.

Movement in Regions of Interest

The video sensor captured a wide view of the subject and their clinical environment, which was used to mark not only *when* movement was detected during recording, but also whether said movement was due to clinical activity, recording artifact, or the subject's own volition. Seven manually marked regions of

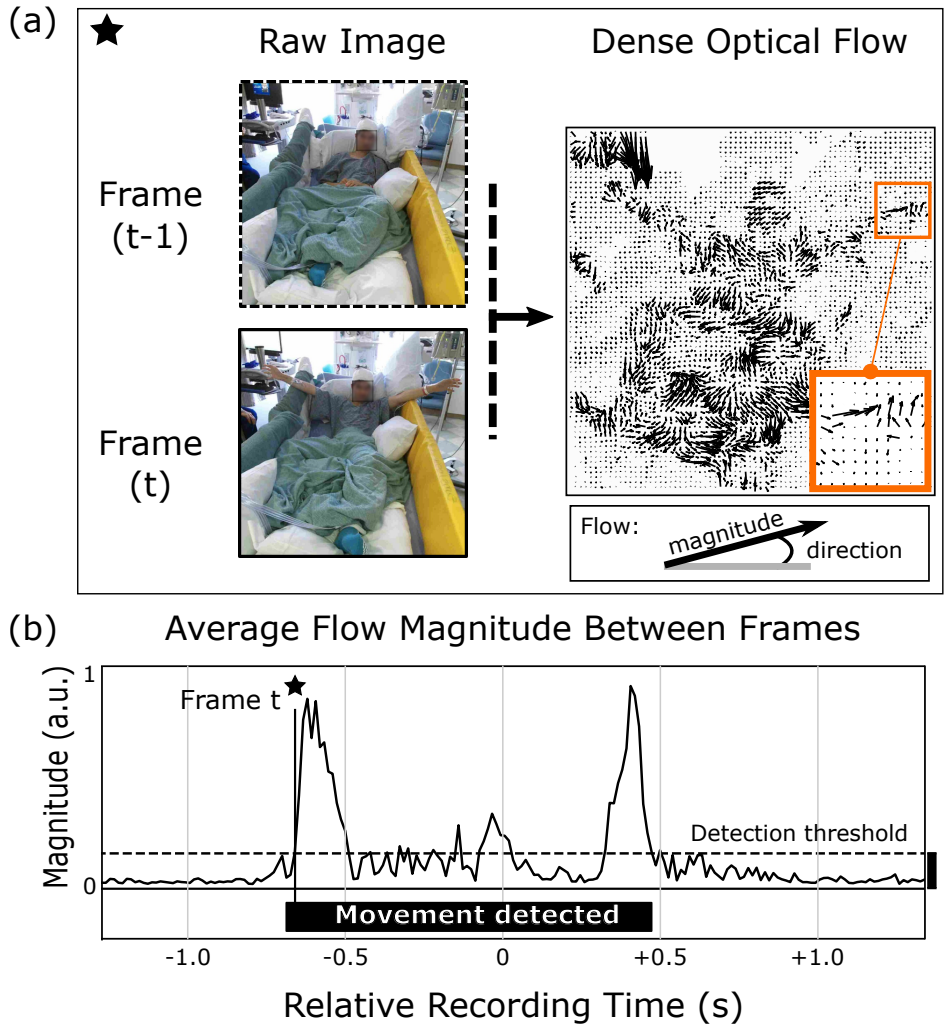


Figure 4.3 **Estimating movement in video using dense optical flow. a)** Exaggerated pair of color frames from the start and middle of an example movement period. The subsequent image displays black arrows representing the dense flow vectors (Farnebäck, 2003) between the two example images. Arrow length and angle represents magnitude [pixels] and direction [radians] of flow observed by each pixel, respectively. For visualization purposes, the flow of every 15^{th} pixel is shown. An expanded view of the subject's right hand moving upwards is highlighted by an orange frame. **b)** Changes in optical flow magnitude over time were used to segment movement periods. Large movements (star) produce large changes in flow magnitude, while continuous movements have tightly-spaced sequences of large magnitude changes.

interest (ROIs) were selected over each video to track periods where the subject was moving ('Head', 'Left', 'Right', 'Body'), along with potential periods of experimental artifacts caused by clinical care (e.g. clinician moving near patient) ('Clinical Boundaries' x2) or camera shaking ('Camera Reference'). Additional descriptions of each region are provided in Supplement 1.3.

Movement, non-movement, and rejection periods during the recording session were automatically marked after thresholding the mean optical flow magnitudes reported from each ROI. These static thresholds

were empirically determined ad hoc from Subject 1 (Table S3) and then applied to every subject. When the signal exceeded threshold, the corresponding flow frames were likely to have represented a sudden spatial change between images. Sustained periods of above-threshold signal were interpreted as movement periods.

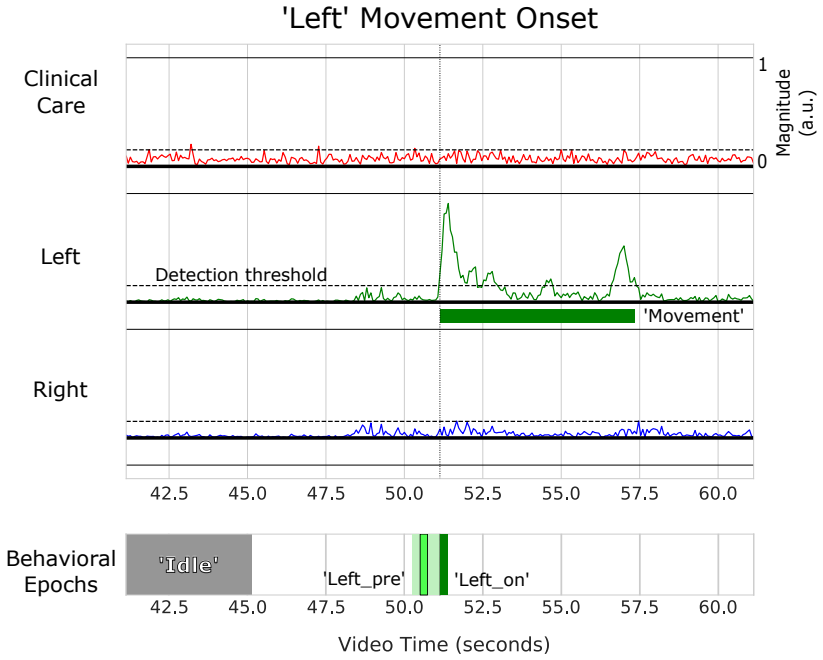


Figure 4.4 **Example of movement onset (Subject 1, 'Left').** A representative example of movement onset is demonstrated over a two second window of average optical flow magnitude, centered around a detected movement period, labeled 'Left'. Movement was detected in the ROI over the subject's left side (green) with no additional activity in other regions, including but not limited to the 'Right' (blue) and 'Clinical Boundaries' (red) ROIs within 150*ms* of the detected onset. Under each of the three signals is the indicator for detected movement in an ROI, previously shown in Figure 4.3. Additional ROIs were used to annotate broader periods when the subject was moving ('Move'), likely to move soon ('Idle', shown in gray), or unlikely to move ('Rest', not shown in black). The final indicator for behavioral epochs shows the time windows from which neural activity would be binned. Specific criteria for behavioral epoch segmentation and selection, including *_pre* and *_on*, can be found in Table 4.2.

Movement Period Labeling

Final movement periods were segmented and labeled from ROI tracking, prioritizing periods where movement was thought to have occurred. First, ROI movement periods were smoothed by merging neighboring periods (within a $2s$ spacing), and then rejecting periods with duration $\leq 1s$. These shorter duration periods were too similar to flow artifacts, which is why they were rejected.

After movement and non-movement periods were separated for each ROI, an additional logic was applied to distinguish 'Left' and 'Right' movement periods that began while no other region detected movement. As a special case, periods when 'Left' and 'Right' movements occurred within 500*ms* of each other

were labeled as 'Joint', which was excluded from study due to a low number of examples.

To contrast these movement labels, 'Idle', 'Rest', and 'Rejected' labels were assigned using a similar logic applied to the remaining periods where no movement was detected across any ROI. Flow frames within $5s$ to $300s$ of movement periods (both onset and offset) were labeled 'Idle', meaning movement was likely to occur but not at that moment. 'Idle' periods spaced within $2s$ were merged together. Then, periods with length $\leq 5s$ were rejected as artifact. Frames further away from any movement periods were considered 'Rest', meaning movement was unlikely to occur. 'Rest' periods within a $2s$ spacing of each other were merged together, and then periods shorter than $\leq 10s$ were rejected.

Several types of time periods were also rejected due to high probabilities of poorly representative data. Periods likely to coincide with clinical activity (e.g. subject seizure, physician care, routine intervention), camera shaking, or movements that require more detailed annotation, were excluded from the study.

Movement Period Review

Finally, video of marked movement periods were reviewed by the authors. Many of these periods identified isolated and intentional movements by the subjects, among more complex activities. Figure 4.4 shows one representative event where movement was only detected over the left side of the subject's body. The marked onset roughly estimates the time when movement signal exceeded detection threshold. For this representative event, one large movement peak is followed by four smaller peaks close enough to be considered as part of the same movement sequence.

4.2.4 Neural Signal Processing

To facilitate comparisons between continuous-monitoring study and more traditional task-based studies, neural features previously identified in structured movement tasks were extracted for unstructured movement epochs/periods. This section describes the steps taken to clean and condition ECoG/sEEG electrode recordings, band-pass filter electrode signals into discrete frequency bands, extract frequency-band power estimates relative to movement labels using an exhaustive grid search, and standardize the extracted features using feature and epoch rejection criteria consistent across all grid search extractions. The pipeline of this process is shown in the right column of Figure 4.2.

Noise Filtering

Both ECoG and sEEG recordings were filtered using second order band-stop filters to remove line noise (60 Hz and harmonics). The resulting time series was high-pass filtered using a 1 Hz cutoff frequency to reduce DC noise.

Electrode Rejection

Each electrode signal was visually inspected by a clinician to reject channels that were excessively perturbed by noise or showed high variance relative to adjacent recordings. Clinical reports were also used to exclude electrodes either near seizure onset zones or exhibiting inter-ictal activity. Additional ECoG and sEEG channels were rejected if the mean squared signal value exceeded a threshold of 3 standard deviations above the mean squared signal value across all channels, indicative of potential noise corruption.

Re-referencing

Electrode channels were re-referenced, either to a common average of all channels to remove shared electromagnetic noise (ECoG) [43, 44] or in a bi-polar [45–47] manner with respect to adjacent electrodes (sEEG).

Band-pass Filtered Power

Spectral power features from the 8–32 Hz (LFB) and 76–100 Hz (HFB) frequency bands were extracted using a windowed finite impulse response (FIR) band-pass filter with linear phase [48]. These frequency bands are the same ones shown to shift during motor repetition tasks by [22]. The power in each band was estimated by calculating the root-mean-squared (RMS) value of a fixed time window relative to the annotated movement periods (Section 2.4.5). The resulting feature matrix produced has shape ($N_{time\_windows}$, $N_{features}$), where the observed features at each epoch is RMS-power of LFB and HFB from each electrode. Band-pass filter details can be found in Table S4.

Feature Selection

This work follows prior demonstrations that informative neural signals can be sampled prior to instructed movement behavior [49, 50]. We hypothesize that neural activity with respect to uninstructed movements will contain information that can be sampled, but the exact window size and offset [31] are unknown. To address this, an exhaustive grid search over two time window parameters – sample window

size and *offset from movement event* – was used to generate candidate datasets of neural activity at different times and scales prior to unstructured movement. Neural window sizes were swept from 200*ms* to 1000*ms* in 200*ms* increments. For each movement event, one window was sampled beginning at movement onset ($*_{on}$) while a second window was sampled prior to marked onset ($*_{pre}$) (Figure 4.4, bottom row). The offset of the latter window was swept from 0*ms* to 400*ms* prior to marked movement onset, in 50*ms* increments. Because non-movement labels are defined by periods of inactivity, baseline 'Rest' and 'Idle' periods were sampled from start to end using non-overlapping bins equal to sample window size.

Epoch Rejection

Any epoch where more than 20% of observed feature values registered as outliers was removed from the entire study. For a given epoch, features were registered as outliers if their observed value exceeded 3.5 standard deviations above the mean of that feature across all epochs.

Dependent Feature Rejection

Confounds from hardware noise (e.g. electrodes shorting) and physical transmission (e.g. cortical volume conductance) may cause neural signal features to be coupled throughout the recording. If this coupling is only present when training a classifier, that classifier will be degenerate if those features are no longer coupled in the test set. Mathematically, tightly coupled features yields rank-deficient inputs that can result in numerical precision and computational errors during testing. To keep features matrices full rank and to reduce computational complexity, linearly dependent features were also removed from the entire study. Such features were identified by sequentially pairing each feature and calculating the difference between their inner products and the product of their norms [51]. Feature pairs were considered linearly dependent if the difference between these two products was below threshold ($\leq 10^{-4}$). Features that were consistently linearly dependent to every other feature were removed from the entire study. Counts of the number of features selected for the study are provided in Table S1.

Balancing Baseline Features across Grid Search

Sampling the same baseline period with smaller window sizes produces more samples from that same period. To avoid biasing towards decreasing window sizes, baseline features at every window size were down-sampled to match the minimum number of features generated from the largest (1000*ms*) window size. Specifically, the features at each window size were matched by nearest time stamp to each baseline

feature generated with the largest (1000*ms*) window. This ensured that baseline features at each window size represented similar points in time relative to the end of their respective sampling windows (Figure S1).

4.2.5 Classification of Movement Onset Labels

We evaluated whether an unstructured paradigm yields separable differences in the neural feature space between coarse movements by training classifiers to distinguish movement onset labels (Table 4.2). Classification paradigms were varied by movement activity, lateralization, and timing; each paradigm was evaluated using three types of classifiers: 1) a linear classifier, 2) a support vector classifier, and 3) a shallow neural network. As shown in Table 4.3, the occurrence of movement onsets were sparse compared to the number of baseline epochs. A nested cross-validation framework [52] was used to account for the finite number of epochs collected for each subject, while evaluating each classifier (Figure 4.5).

Classifiers

The following three models were used for unstructured movement classification:

- Linear Discriminant Analysis (LDA)
- Support Vector Machine (SVM)
- Multilayer Perceptron (MLP)

These classifiers were used to evaluate whether data collected in this experiment were informative. Various models were selected to measure the degree of non-linearity required to separate neural features by movement type, if at all possible. Additional model details can be found in Supplement 1.4.

Classification Paradigms

Models were evaluated on the following five multi-class paradigms:

- 'Rest' vs 'Idle'
- 'Idle' vs 'Move$_{pre}$' vs 'Move$_{on}$'
- 'Left$_{all}$' vs 'Right$_{all}$'
- 'Left$_{pre}$' vs 'Right$_{pre}$'
- 'Right$_{pre}$' vs 'Right$_{on}$'

The paradigms selected correspond to specific assessments of the neural features collected under unstructured conditions. 'Rest' vs 'Idle' compared conditions where the probability of the subject moving was respectively low or high. 'Idle' vs 'Move_{pre}' vs 'Move_{on}' emulated the three stages of movement initiation, starting from non-movement and ending on movement onset. The coarse labeling minimized any potential overlap of pre-movement and non-movement activity, but also dropped any relationships between paired '$_{pre}$' and '$_{on}$' samples. The 'Left' vs 'Right' paradigms compared the nuances of movement conditions by which side moved. Finally, 'Right_{pre}' vs 'Right_{on}' compared neural features for extracted from two time windows relative to only right-sided movements. The right side was considered first because each subject had electrode implants on their left hemispheres, and neural response to motor activation has been shown to be predominantly contralateral [53–55]. Table 4.2 further defines each movement onset class.

Model Training and Evaluation

A nested cross-validation framework was used to evaluate each classifier while reducing the bias from re-sampling an already limited dataset [52]. As shown in Figure 4.5, the nested framework consisted of an outer 5-fold cross-validation layer for *classifier evaluation* and an inner 5-fold cross-validation layer for *classifier optimization*. The inner layer was used to select the optimal model based solely on training data, which was then evaluated on the test data held out in the outer layer. The nested framework produced estimates of the classification performance, tested five times, of an optimized classifier, tested an additional five times. For each fold, features for both train and test data were z-scored with respect to the distribution of the training set cropped for the given classification paradigm.

As shown in Table 4.3, the number of examples for each movement class are unbalanced and overly represent non-movement epochs. A naive classifier trained on as much data as possible, if unbalanced, typically learns to always predict the majority class. To address this during training, the weighting of each class was adjusted by one of two ways. For the LDA classifiers, the entire training set was used while setting the prior probabilities of each class in the classifier to be uniform. Such an option was not available for the SVM and MLP classifiers, so training data was resampled to represent each class as many times as the majority class, capped at N=125 samples per class. This meant that for a given dataset, each class was re-sampled so that the prior probability of sampling every class in the new dataset was equal.

During evaluation on test data, as much data as possible was fed into the trained models. The predicted labels were compared to ground truth on a class-by-class basis, in order to produce unweighted measures of model performance. Classification accuracy averaged across folds is the primary metric reported in this work, but we also provide sensitivity and specificity reports, as well as confusion matricies in

Table 4.2 Description of conditions assumed for each movement onset class label.

Label	Description
'Rest'	No movement was detected in any monitored region for at least 300*s*, so the probability of movement is low. Periods with this label are at least 20*s* in duration.
'Idle'	No movement was detected in any monitored region for at least 5*s* but no more than 300*s*, so the probability of movement remains high. Periods with this label are at least 5*s* in duration.
'Move'	Collection of 'Left', 'Right', and 'Joint' labels to represent movement onsets in general.
'Left'	Movement onset detected over the left side of subject's body only. Periods with this label are at least 1s in duration.
'Right'	Movement onset detected over the right side of the subject's body only. Periods with this label are at least 1s in duration.
'$*_{pre}$'	Pre-movement subscript denotes that neural activity was sampled prior to marked movement onset.
'$*_{on}$'	During-movement subscript denotes that neural activity was sampled starting at marked movement onset.
'$*_{all}$'	All-inclusive subscript denotes union of '$*_{pre}$' and '$*_{on}$'.

the supplement (Tables S6-10, Figures S5-7,9-11).

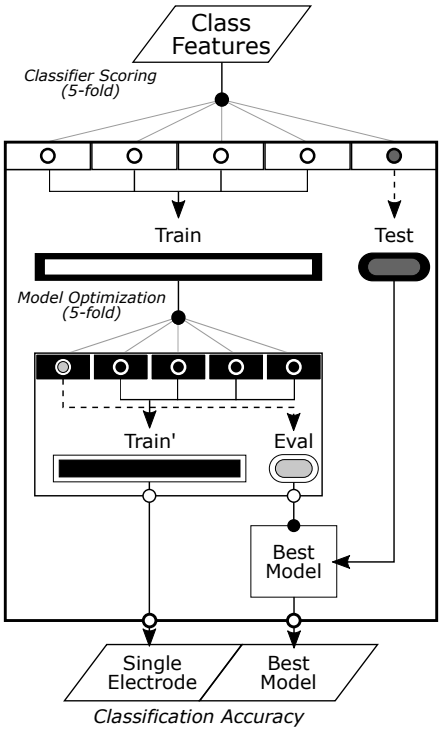


Figure 4.5 **Optimizing and evaluating movement onset classifier models.** Nested cross-validation was used to optimize and evaluate classification models while considering the limited sample counts of the (sparse relative to baseline) movement onset labels. Framework consisted of an outer 5-fold cross-validation layer for *classifier evaluation* and an inner 5-fold cross-validation layer for *classifier optimization*. The inner layer was used to select the optimal model based solely on training data, which was then evaluated on the test data held out in the outer layer. The steps shown were applied to each window size/offset candidate of neural features.

Classifier Optimization and Selection

To select the best version of a classifier for a given training set, classifier hyperparameters were optimized in the inner cross-validation layer. LDA classifiers were optimized according to the number of features selected for the model. Candidate LDA classifiers were trained by incrementally adding both LFB and HFB features from individual electrodes. Electrodes were ordered by their individual classification accuracy, which was measured using similar LDA classifiers. A single SVM classifier was trained per outer fold of the nested cross-validation framework, using the default parameters of scikit-learn. While one advantage of MLP classifiers is that the hidden layers can assign weights to all dataset features, the exact shape and weighting of the MLP framework introduces an additional set of hyperparameters. An exhaustive grid search was used to sweep number of hidden layers and nodes per layer to generate candidate MLP classifiers. The candidate model with the highest validation accuracy was then evaluated on the test data.

The exact parameters considered for classifier construction and training can be found in Supplement

Table 4.3 Comprehensive list of coarse movement classes considered in the study, the total number of trials generated by subject, and the percent of total trials (%) by label. 'Idle' and 'Rest' trials were generated by binning across entire label durations, while 'Left', 'Right', 'Joint', and 'Move' trials were generated with respect to the movement onset. Due to low sample count, 'Joint' labels were excluded from study. For exact counts rather than percentages, please refer to Table S5.

Subject No.	Total Samples	Idle (%)	Rest (%)	Left (%)	Right (%)	Joint (%)	Move (%)
1	2812	85.4	5.0	2.4	1.8	0.6	4.8
2	18662	52.1	41.5	1.2	1.5	0.5	3.2
3	43757	40.4	56.5	0.5	0.9	0.2	1.5

1.4.

Classification Baseline

Sample sizes of this study were limited (as is generally true in BCI studies), so two tiers of theoretical chance levels served as baselines against classifier accuracies. The less stringent 'infinite'-level of chance assumes an infinite number of samples are theoretically available for measuring classifier performance. Because classes were equally represented in the test sets, the 'infinite'-level of chance was equal to $(N_{classes\_tested})^{-1}$. However, the number of trials per class in this study ranged from 10's to 1000's, which is far from infinite, motivating the more stringent 'finite'-level of chance. This theoretical chance level is calculated based on the statistically significant ($p \leq 0.05$) classification performance of a mock dataset randomly drawn from a binomial distribution whose sample size and class counts equaled the real test set [56]. The 'finite'-level baseline was used to consider the possibility of randomly achieving above class performance for limited dataset sizes. Classifiers with test accuracies above the 'finite' baseline were considered statistically significant. The exact values of these baselines and their parameters are provided in Table S10.

4.3 Results

The following sections describe products of movement segmentation, feature extraction, model evaluation, and electrode scoring produced by the previously described methods. Summaries of this un-

structured dataset by subject, label, and overall count can be referenced in Tables 4.1-4.3.

4.3.1 Semi-automated Annotation of Unstructured Movement Onset

Dense optical flow was used to semi-automatically annotate unstructured movement periods for each subject (Figure 4.6, Subject 1). Manually labeled ROIs (Figure 4.6a) were used to aggregate dense optical flow as signals for motion in each region over time (Figure 4.6b). Each signal was then segmented into the movement periods (Figure 4.6c) and consolidated into coarse movement labels. Besides tracking subject-centered movements, it was necessary to monitor movements caused by clinical care. As shown in Figure 4.6c, artifacts caused by clinical interference were distributed throughout the recording session, and had to be seriously accounted for.

4.3.2 Neural Correlates to Unstructured Movement Onset

Figure 4.11 shows the top electrode with noticeable separability in neural features between *all* 'Left' (green) and 'Right' (blue) behavior epochs. The distributions are of LFB and HFB RMS-power, sampled from a window size of 200*ms*, and taken 350*ms* prior to movement onset. Similar to task-based movement studies, we identified behaviorally modulated neural features, including a characteristic [22] power increase in HFB with an accompanying decrease of power in LFB (red 'X'). This top electrode was identified in Subject 1 located over the superior aspect of the post central gyrus, part of somatosensory cortex (S1), of the left hemisphere. Placement on the left hemisphere may also explain why the changes in neural spectral power are more pronounced for the contralateral movement label, 'Right', over the ipsilateral movement label, 'Left'.

Spectral features from all examples of each class are included in the figure. Plotting data was normalized after power computation, using all examples from each class being compared. Similar plots of feature distributions for Subjects 2 and 3 are provided in Figure S2.

4.3.3 Feature Response Mapping

The results of single-electrode evaluation using LDA are shown in Figure 4.8. Each electrode was scored by the ability of their features (HFB, LFB) to correctly classify movement onset labels, such as 'Left_{all}' vs 'Right_{all}'. Specifically for sEEG, the 'electrode' refers to the virtual electrode that results from bi-polar re-referencing. The first column shows individual electrode scores averaged across 5-folds cross-validation, showing which electrodes scored above the less stringent 'infinite' chance level using a divergent

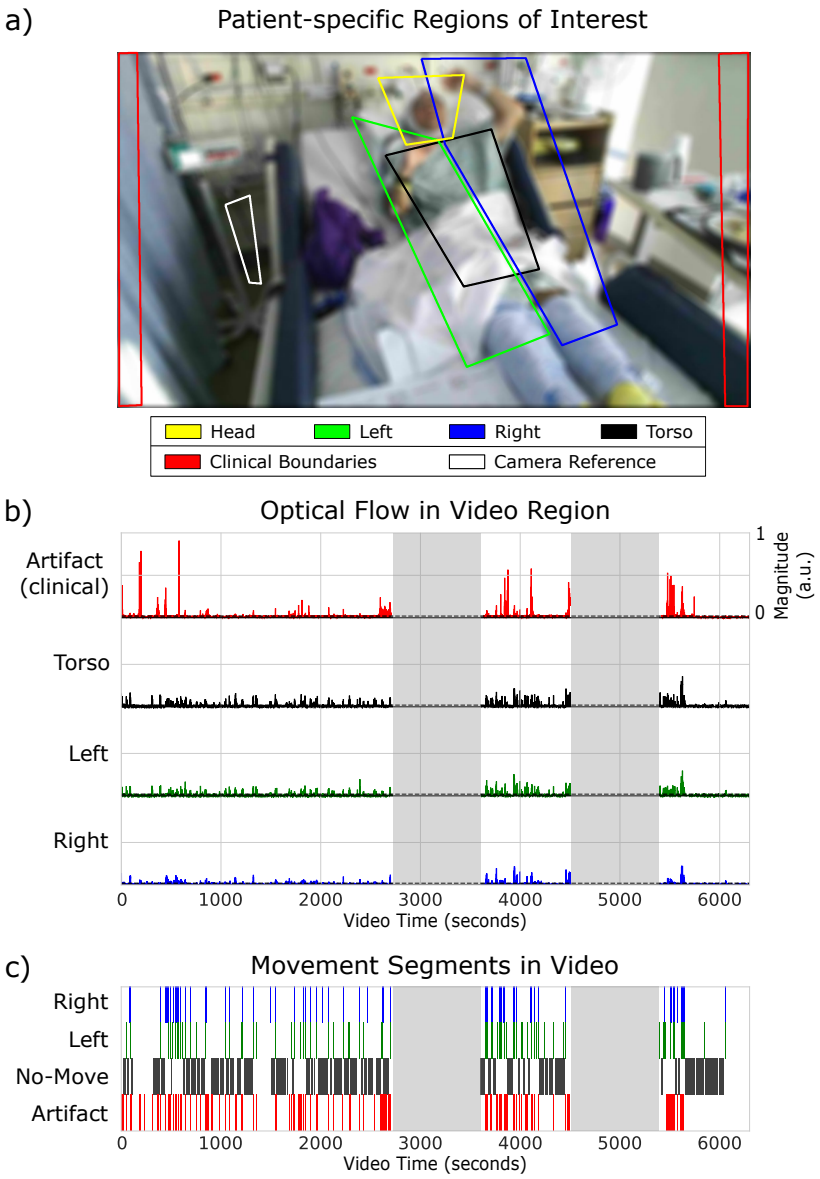


Figure 4.6 **Video segmentation via region of interest (ROI) tracking of optical flow magnitude (Subject 1).** Annotations of movement and non-movement periods were generated by thresholding the magnitude of optical flow in manually defined ROIs. **a)** An example frame with ROIs manually marked for appearance of clinical care (red), motion artifact (white), and patient movement (various). **b)** Stacked time series of the average magnitude of flow within each ROI. Regions in frames with flow above 0.35 pixels/frame were marked as movement detected. The magnitudes are visualized in unit scale. **c)** Detected period of movement were marked with a start and end time, as well as a coarse behavioral label. Regions of non-movement were also marked to serve as a baseline for movement onset classification. In both **b)** and **c)**, time periods where view of the subject was covered were excluded (grey).

color map. In the second column, electrodes were grouped into three discrete categories based on how their score compared to both 'infinite' and 'finite' levels of chance. (pink) Electrodes with a solo classification score above the less stringent 'infinite'-level of chance, but below the more stringent 'finite'-level of chance,

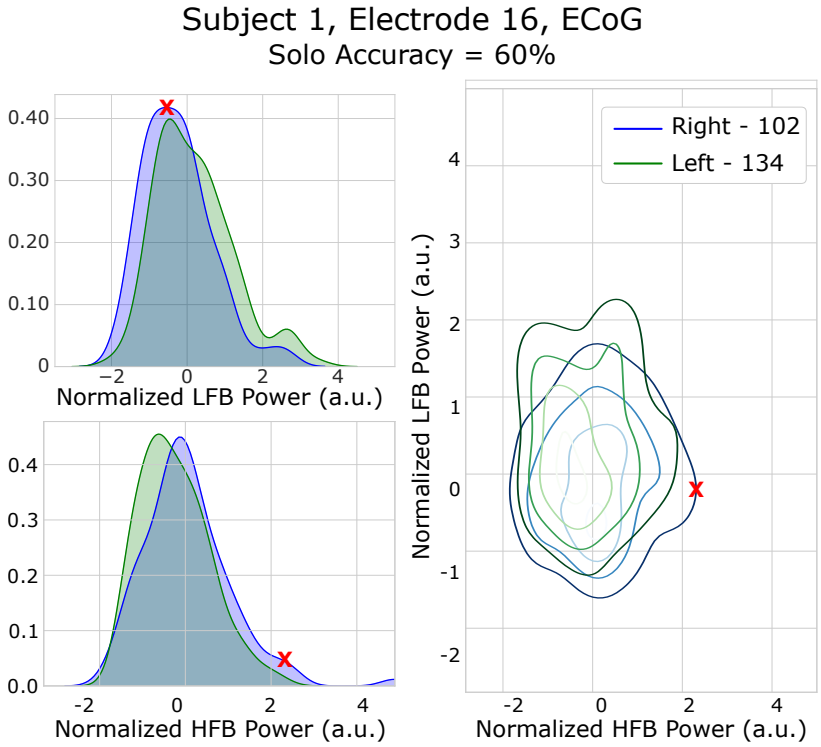


Figure 4.7 **Example of neural features from top electrode over sensorimotor region (Subject 1).** Distributions of 8-32Hz low frequency band (LFB) and 76-100Hz high frequency band (HFB) RMS-power, sampled from a window size of 200*ms*, and taken 350*ms* prior to movement onset. We identified behaviorally modulated neural features, including a characteristic power increase in HFB with an accompanying decrease of power in LFB (red 'X'). This top electrode was identified in Subject 1 located over the superior aspect of the post central gyrus, part of somatosensory cortex (S1), of the left hemisphere. Being located on the left hemisphere may also explain why the changes in neural spectral power are more pronounced for the contralateral movement label, 'Right' (blue), over the ipsilateral movement label, 'Left' (green). The total number of epochs for each class are reported in the legend.

suggest marginal distinction between features; (yellow) electrodes with a solo classification score above the 'finite'-level of chance suggest linearly distinguishable distinctions in estimated band power between movement labels; (blue) electrodes with solo classification below both baselines suggest no distinction between labels, 'Left_{all}' vs 'Right_{all}'. For all subjects, these electrodes were identified over or near primary motor cortex (M1) and S1, as shown in the post-implant electrode reconstruction maps for each subject. These maps also identified significant electrodes near middle temporal gyrus (MTG). Similar maps using HFB and LFB features separately are provided in the Supplement Figures S3,4.

4.3.4 Movement Onset Classification

Figure 4.9 summarizes the best model scores (mean, SEM) for each classification paradigm, classifier, and subject following nested 5x5-fold cross-validation (Figure 4.5). When comparing classifier

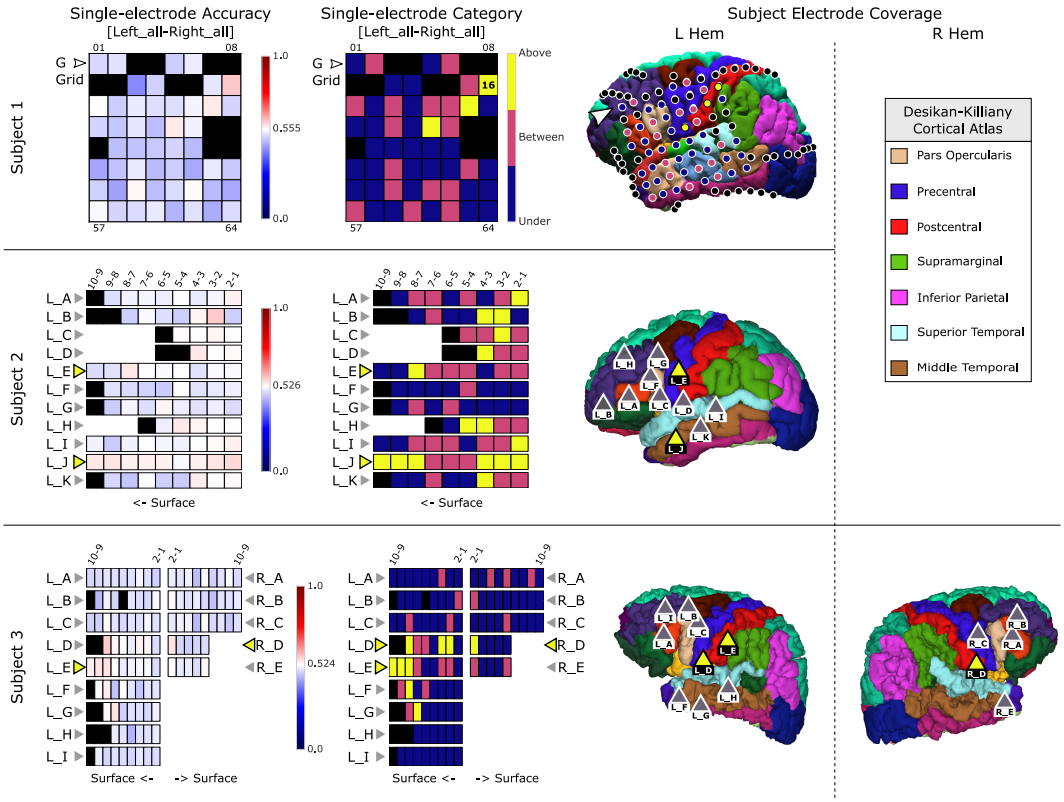


Figure 4.8 **Mapping single-electrode scores for 'Left$_{all}$' vs 'Right$_{all}$' onset classification.** Each electrode was scored by the ability of their features (HFB, LFB) to correctly classify movement onset labels, such as 'Left$_{all}$' vs 'Right$_{all}$'. The first column shows each electrode score averaged across 5-folds cross-validation, showing which electrodes scored above (red) the less stringent 'infinite' chance level using a divergent color map. Electrode references are also provided to orient onto each subject's electrode reconstruction map (third, fourth column). In the second column, electrodes were grouped into three discrete categories based on how their score compared to both 'infinite' and 'finite' levels of chance. (pink) Electrodes with a solo classification score above the less stringent 'infinite'-level of chance, but below the more stringent 'finite'-level of chance, suggest marginal distinction between features; (yellow) electrodes with a solo classification score above the 'finite'-level of chance suggest linearly distinguishable distinctions in estimated band power between movement labels; (blue) electrodes with solo classification below both baselines suggest no distinction between labels, 'Left$_{all}$' vs 'Right$_{all}$'. For all subjects, these electrodes were identified over or near central sulcus. The third and fourth columns show the post-implant electrode reconstruction maps for each subject. Subjects with ECoG coverage (Subject 1) show electrode responses mapped to the cortical surface. Subjects with sEEG coverage (Subjects 2,3) show where along the cortical surface each probe was inserted.

performance to baseline, most classifiers exceeded 'infinite' and 'finite' levels of chance for each subject and movement condition. We note that in several cases such as (Subject 2, '*Idle*' vs '*Move$_{pre}$*' vs '*Move$_{on}$*', MLP classifier), performance is near 'finite' levels of chance classification. The scope of model and feature selection parameters was limited due to the overall complexity of neural features likely underlying coarsely defined labels. Nonetheless, the above-chance performance of both linear and nonlinear classifier mod-

els suggests that we successfully collected informative neural activity with respect to unstructured motor behaviors. For a report of the corresponding confusion matrices, see Figures S3-5.

As a reminder, feature selection was only optimized for the LDA models (Figure S8). This was because the linear decision boundaries of LDA were more susceptible to poor fit when compared to SVM and MLP models. SVM and MLP classifier accuracies are likely to be improved with feature selection and/or dimensionality reduction techniques, but requires further study.

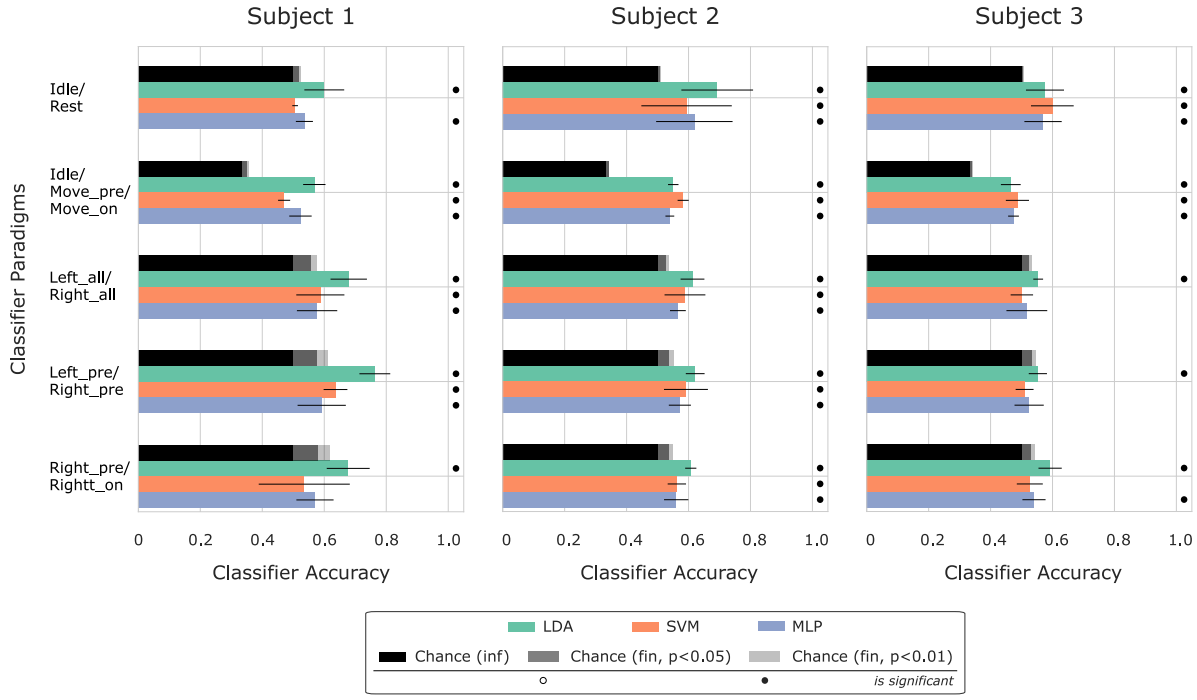


Figure 4.9 **Best model scores for each classifier paradigm and subject.** Unbalanced Linear Discriminant Analysis (LDA), balanced Support Vector Machine (SVM), and balanced Multi-layer Perceptron (MLP) classifiers were trained on features from movement ('Left', 'Right', or 'Move') and non-movement periods ('Idle', 'Rest'). Uniform class priors were enforced while training each model, while test set results were balanced by class to allow comparison between classifier paradigms. Classifier accuracies are reported as bar graphs that summarize model evaluation over 5-fold cross validation. Classifiers with average accuracy above 'finite' chance are marked.

4.4 Discussion

This work describes methods for annotating movement behaviors from continuously recorded video, quasi-trialization and extraction of neural features with respect to these movement behaviors, and rigorous evaluation of whether neural activity can classify movement behaviors. These methods also produced subject motion reports, heat maps of feature response to different movement modalities, and classifier

results that were significantly above chance for all subjects and classifier paradigms. The results of this work serve as a preliminary demonstration that recording and studying continuous, unstructured behaviors can be informative for studying neural activity not evoked by tasks, which may help generalize future neural decoding algorithms. The assumptions, challenges, and limits of this work are discussed below.

4.4.1 Unstructured Movement

Several types of movement modalities were observed during recording. Namely, the movements detected for each subject included reaching, scratching, and playing with electronics like television controller or cellphone. Example videos of subject movements are included in Supplement V1-3, showing that the detected movements included in this study were primarily reaching and scratching. A legend for these videos is provided in Figure S9.

4.4.2 Caveats of Unstructured Movement

By design, continuous monitoring aims to capture a much greater variety of unstructured behaviors when compared to structured experimental tasks. This is done by recording subject behavior over longer periods than a typical task-based session, which are on the order of hours and minutes, respectively. The resulting data from continuous monitoring was long and unlabeled. Without task-driven cues or markers, new behavioral labels were constructed. These labels coarsely grouped movement conditions by whether they occurred over the subject's left or right side.

Coarse categorization of movement behaviors ignored many of the nuances between individual movements and cortical activity, yet we believe that the results of classifying movement onset periods by type may still yield meaningful insight into the generalization of neural decoding algorithms. As demonstrated in [33], neural activity on larger time scales (minutes) may be informative for decoding broader behavioral context. While the use of coarse behavioral labels yields higher counts of examples per class, the variance between individual movement epochs also increases. These nuances may be accounted for with more detailed characterization of subject movement [57–59], and future work is being pursued in this regard.

Table 4.3 shows the total number of movement onset epochs extracted for each subject. A majority of such epochs belong to either 'Rest' or 'Idle', in part due to the sedated nature of individuals in the EMU. Because this work focuses on motor behaviors, we aimed to record the subject when they were most active. This meant recording during periods of the day when the subject was awake, engaged, and moving. Compared to known trial-based experiments [60] whose trial counts are on the order of tens per behavior,

the number of samples collected in this experiment is on the order of hundreds. The exception is Subject 1, whose movement trial counts are also on the order of tens of samples, as only 1.5 hours of their recorded data were included. In other words, the total trial counts collected under a continuous paradigm greatly exceeded the trial counts collected under a typical task-based paradigm, primarily due to increased recording time.

To summarize, coarsely labeling unstructured behaviors as done in this study yields feature sets that are highly variable, broadly defined, and have an overwhelming representation of non-task-related activity. Taking these caveats into consideration, nested cross-validation was introduced as a rigorous framework for tuning and evaluating each experiment consisting of a set of neural features, classes being compared, and type of classifier. Furthermore, two baseline criteria for chance performance were used to provide more stringent comparison of classifier accuracies, taking into consideration the finite number of epochs included. These evaluations, particularly the classifier accuracies, are likely still constrained by sample size and could be improved by extending recording to the full extent of patient stay in the clinic. In addition, better labels and metrics for annotating movement behaviors could be designed to fully leverage subject recordings. Such ideas have been considered in [33, 57], and serves as future directions for the work presented in this paper.

4.4.3 Classification of Movement Onset

We believe to have captured neural activity modulated by unstructured motor behaviors. To demonstrate this, a set of classifiers was trained on neural features, previously identified from [22], to distinguish between coarse movement conditions. These conditions represented a few things. 'Rest', 'Idle', and 'Move' labels distinguish periods of time where movement was unlikely to occur, movement may have occurred, and movement was about to/in the process of occurring, respectively. Furthermore, we inspected the separability between types of movement behaviors, namely lateralization ('Left' vs 'Right') and timing ('$*_{pre}$' vs '$*_{on}$'). The classification paradigms reported in the manuscript were selected to represent possible state comparisons that future motor BCIs may face – off/idle, stop/go, move left/move right. These movement contexts may serve as elements of basic motor BCI functionality, though additional work is needed to explore this space, particularly when classifying more behaviors (Figures S10-12). The positive performance of both linear and nonlinear decision boundaries motivates further exploration through machine learning and neural signal processing algorithms to better parameterize neural activity with respect to unstructured movements. Nonetheless, the initial findings suggest that these volitional behaviors may be separable in the neural feature space.

When comparing results between subjects, note that the coverage of electrodes around sensorimotor cortex (SMC) is inconsistent. Previous motor physiology works like [49, 50] identify strong correlations

between movement tasks and modulation in SMC, making it an important cortical area from which to record. For example, Subject 1 had reasonably wide coverage across SMC from ECoG, specifically ventral primary motor cortex (M1v) and S1 regions, and this was reflected in the classification performances as shown in Table S6 and Figures 9, S3, and S7, whereas the coverage of sEEG in Subjects 2 and 3 was not as sufficient to well distinguish left- and right-sided movements at this coarse level of detail. The maps of single-electrode scores shown in Figures 8, S3, and S4 help identify which individual electrodes are informative of movement behaviors, and the differences in how LFB and HFB features contribute as well. For the ECoG subject (S1), electrodes with significant scores using LFB were distributed over M1 and S1 regions, whereas one electrode over S1 was significant using HFB. For the sEEG subjects (S2,3), there appeared to be clustering of electrodes with significant scores using both LFB and HFB separately, in electrode shanks implanted near S1, M1, and MTG. We note that several of these electrode pairs were over white matter of frontal and temporal lobes. Signals from white matter can result from volume conductance (zero phase) of activity from nearby gray matter, but can also contain white matter specific signaling [61]. The quality of these signals differ depending on their re-ferencing [44], with methods like bi-polar or laplacian able to account for differences in cortical matter with some success. The reported significance of these signals may also by affected by other factors such as overlapping behavior contexts, so definitive conclusions about origin of these significant signals require further exploration. While our preliminary work shows that individual electrodes have informative features, further work with more samples and subjects is required to identify interpretable correlation between behaviors and individual features.

4.4.4 Implications and Applications

Part of the methods proposed in this work generate annotations of video and neural data taken during periods of clinical care. Motion reports summarized not only how much the subject moved, but also tracked periods of possible clinical care, as part of monitoring for possible experimental artifacts interacting with the subject. If applied in real time, these reports could augment the monitoring of subject-specific physical activity. Electrode evaluation produced maps identifying electrodes that were informative for classifying coarse movement types.

In regard to unstructured movement, the electrode maps from Figure 4.8 tended to identify informative electrodes over M1 and S1, which is consistent with previous motor physiology work [49, 50]. These maps also identified significant electrodes near MTG. Due to the limited subject count and clinically driven placement of electrodes in this study, it is difficult to draw specific conclusions about the regions over which these top electrodes were identified. With further work, such maps could be used to augment clinical func-

tional mapping, and even surgical planning. Providing additional confidence can inform surgical decision making by possibly identifying critical functional areas near implanted electrodes.

Further work and the advancement of the methods described here could enable related studies of continuous intracranial behavior, including but not limited to the following: 24/7 resting-states and their neural correlates, neural population dynamics of consistent behaviors, automatic labeling of epilepsy data, or the influence of medication on the neuronal data. To study these possible applications will require robust day over day consistency in results for the same subject, which is an open topic of interest that we believe can be addresses when more data is collected. A few works have shown that decoding of neural activity between days is possible but with some degradation, both in monkeys and humans. [32] shows that ECoG signal can be coherently clustered into broad speech and movement labels across 4 days of recording. [62] explores this question from a neural prosthetics perspective, demonstrating that while BCI performance drops further away in time from training, there are methods to improve stability across days of usage.

4.4.5 Comparison to Previous Work

The work presented in this paper is not the first to study neural correlates to unstructured behaviors [7, 25, 31, 33] or long-term continuously recorded EMU data [30, 32]. This work does, however, serve as a proof-of-principle that a unstructured movement behaviors that vary in magnitude and lateralization can be identified and labeled from continuously recorded data, and that neural activity with respect to these labels is informative. In terms of experimental design and classification performance, our study is most similar to the work done by [31]. The aforementioned work specifically focuses on predicting movement of the left wrist, tracking movement of the limb using a pose tracking computer vision algorithm. Their predictive model relied on a deep learning framework to engineer features from raw neural signal, and was able to predict with 60-70% accuracy if the left wrist is moving or at rest based upon video and neural activity from a fixed window prior to the known movement onset. Similarly, we demonstrate classification with 60% accuracy of movement onset after observing a sequence of video and neural activity prior to known movement onset. In contrast to the previous study, we show significant accuracy for classifying which side is moving, as well as whether the movement is about to or has begun. Our work uses conventional spectral power features and compares multiple models, in part motivated by the coarse behavior labels annotated using computer vision optical flow. Combined, these works demonstrate the viability of and provides methods for studying unstructured motor behaviors in the EMU – specifically that coarse movement labels and more precise pose tracking annotations can be studied with neural activity. Our unique contributions and improvements to this field present the first use of dense optical flow to semi-automatically detect and label

visible behaviors (including clinical care and left/right sided movement); this work is also the first to assess the performance of conventional neural features learned in task-based movement studies when classifying similar but unstructured movements.

4.5 Conclusion

This work demonstrates that neural correlates to uninstructed behaviors, specifically unstructured movements, can be collected, scored, and evaluated using our proposed methods. This was accomplished by using coarse labels to group unstructured motor behaviors, focusing on neural activity sampled near the onset of these behaviors, and applying a rigorous evaluation framework to classify behavioral labels using spectral features of said neural activity. A system was built to collect hours of such patient data. Semi-automated scripts were used to annotate patient movement periods and to extract neural spectral features with respect to movement annotations. Neural features from various movement conditions were used to train and evaluate classifiers to imitate future motor BCI paradigms in unstructured contexts. These methods can be applied to other similar recording paradigms and continue the motivation for continuous monitoring in the clinic. The multi-modal datasets collected through this work motivate future studies of neural correlates to more unstructured behaviors, such as movement kinematics or behavioral contexts for subjects in the EMU.

Author Contributions

The nature of BCI research is highly collaborative by design and requires contributions from experts across the fields of clinical neurology, neuroscience, and engineering. To enable this work, the authors contributed in the following ways: Electrode implantation was performed via surgery by WKD and DG. Clinical monitoring of the epileptic subjects was provided by PD, DF, OD, SS, and SGW. Data acquisition was coordinated by LM, TT, PGG, and advised by VG. Behavior labels were annotated by PGG, KJC, and AA, advised by VG. MLP architectures were created by PGG, assisted by TP, and advised by VG. Analysis and writing were led by PGG, assisted by AA, and advised by VG.

Acknowledgements

This work has been supported in part by the UCSD ECE Department Medical Devices & Systems Initiative, the UCSD Centers for Human Brain Activity Mapping (CHBAM) and Brain Activity Mapping

(CBAM), the UCSD Frontiers of Innovation Scholars Program, the Qualcomm Institute Calit2 Strategic Research Opportunities (CSRO) program, the Hellman Fellowship, the Institute of Engineering in Medicine Graduate Student Fellowship, and the Clinical and Translational Research Institute at UC San Diego.

We would like to thank those at UC San Diego, Rady Children's Hospital of San Diego, and the Comprehensive Epilepsy Center at NYU Langone Medical Center. We specifically thank Chenghao Gong, Preet Minas, and Hugh Wang for their contributions to data collection and annotation.

This chapter is a reprint of the material as it appears in Gabriel, Paolo Gutierrez, Kenny Chen, Abdulwahab Alasfour, Tejaswy Pailla, Werner Doyle, Orrin Devinsky, Daniel Friedman, Patricia Dugan, Lucia Melloni, Thomas Thesen, David Gonda, Shifteh Sattar, Sonya Wang, and Vikash Gilja. "Neural correlates of unstructured motor behaviors." *Journal of neural engineering* (2019)..

Code Availability

The tools and code used in this work are available upon request to the corresponding author.

4.6 Supplement

4.6.1 Intracranial Electrodes

The technical aspects of both electrode types used during recording are described below:

Electrocorticography (ECoG) Electrodes were embedded in SILASTIC sheets (2.3 mm exposed diameter, 10 mm center-to-center spacing), and manually placed directly on the cortical surface. ECoG activity was recorded from 0.1 to 230 Hz using Nicolet clinical amplifiers, and the data was digitized and sampled at 512 Hz while referenced to a screw bolted to the skull at a cm distance from the craniotomy edge under the scalp.

Stereotactic electroencephalography (sEEG) Electrode contacts had a 0.8 mm diameter, 2 mm length, and 3-5 mm center-to-center spacing. sEEG electrodes were inserted via the ROSA robotic surgical implantation (Medtech Surgical Inc., USA) in various orientations to allow for intracranial recording from lateral, intermediate, and/or deep cortical and subcortical structures in a 3-D arrangement. sEEG data was sampled at 2000 Hz using Xltek 128Fs clinical amplifiers. For each subject, one sEEG electrode contact far removed from the seizure onset zone was used as a reference electrode, as bi-modal re-referencing is standard practice for the clinicians.

4.6.2 Signal Alignment

Kinect and neural recording hardware were time-aligned using an Arduino microcontroller to periodically (T=120s) send a random number (1:8) of characteristic pulses to both systems. The neural system received the signal as a series of square waves, whereas the Kinect microphones recorded the signal as a series of tones. The onset and number of pulses at each sync period was extracted from each system and matched to align both data streams.

Due to noise in the audio recording, there was uncertainty in the onsets extracted from the Kinect data streams. This jitter was manually corrected after extraction using a custom Python script. With this intervention, recording alignment was constrained to 5ms, which is within Kinect video sampling resolution.

4.6.3 Region of Interest Definitions

Seven regions of interest were manually marked over each subject video to track periods where the subject was moving during recording and potential periods of experimental artifact caused by clinical care or camera shaking. They are defined as follows:

- 'Clinical Boundaries' (red) - defined on the left and right sides of the subject to identify possible periods of clinical artifact, such as a nurse providing standard care, visitors interacting near the subject, or a clinician conducting a physical assessment.
- 'Camera Reference' (white) - defined over a fixed landmark in the clinical space, such that movements detected in this region likely meant that the camera was moving.
- 'Head' (yellow) - defined over the subject's head as a *supplementary* marker for potential periods where electrode signals were made noisy due to iEEG cable movements.
- 'Left' (green) - defined over the left side of the subject's body to monitor movements of their left upper body area.
- 'Right' (blue) - defined over the right side of the subject's body to monitor movements of their right upper body area.
- 'Body' (black) - defined over the subject's body to monitor movements of their upper body area.

4.6.4 Classifiers - Details

Linear Discriminant Analysis (LDA)

To test whether movement labels could be separated with linear decision boundaries, a Linear Discriminant Analysis (LDA) classifier was used (scikit-learn, *LinearDiscriminantAnalysis*). This method uses Bayes' rule to fit Gaussian prototypes of each class to the data. When initialized, classifiers of this type were set with uniform class prior probabilities and covariance of each class prototype were shared. During training, eigenvalue decomposition was used as the solver, and regularization was applied by forcing the shared covariance matrix to be diagonal.

Support Vector Machine (SVM)

To test whether movement label classification would benefit from non-linear decision boundaries, a Support Vector Machine (SVM) classifier was used (scikit-learn, *svm.SVC*). A 'rbf' kernel was used. Shrinkage was applied during training. Class weights were balanced, as the training set was balanced prior to training.

Multilayer Perceptron (MLP)

To further test whether movement label classification would benefit from non-linear decision boundaries, a Multi-layer Perceptron (MLP) neural network was used (scikit-learn, $neural\_network.MLPClassifier$). A few shallow MLP architectures were generated by varying the 5-50 nodes in 1-2 hidden layers, each with rectified linear unit activation function. Each MLP classifiers was trained using a batch size of 10 samples, a constant learning rate (0.001), and early stopping.

4.7 Supplementary Tables

Table 4.4 Electrode and feature counts for each subject. Over 100 electrodes were implanted for each subject, but some channels were excluded from study, as described in the manuscript. Additional rejection was applied after extracting spectral power features from each channel. The percentage of features that were found to be linear, and subsequently were rejected, are included in parentheses.

Subject No.	Electrodes included / implanted (count)	Features selected (count, % linear)
1	51 / 118	102 (0)
2	89 / 127	156 (12.3)
3	121 / 128	214 (11.5)

Table 4.5 Summary of total video recording time, camera frame rate (automatically set by the sensor), and total number of frames processed for each subject.

Subject No.	Recording duration (hrs.)	Frame rate (fps)	Processed frames (count)
1	2.25	15 or 30	67472
2	9	30	571130
3	24	30	1538492

Table 4.6 The fixed thresholds applied to each ROI signal.

ROI Category	Threshold (pixels per frame)
Subject	0.35
Reference	0.5
Clinical Care	1.0

Table 4.7 Filter details for the windowed FIR band-pass filters used to extract LFB (8-32Hz) and HFB (76-100Hz) features. Filters were designed using NeuroDSP by (Cole et al., 2019).

Subject No.	Sampling Rate (Hz)	Phase	Cutoff Frequencies (Hz)	Order
1	512	linear	8 – 32	193
			76 – 110	21
2	2000	linear	8 – 32	751
			76 – 110	79
3	2000	linear	8 – 32	751
			76 – 110	79

Table 4.8 Comprehensive list of coarse movement classes considered in the study, as well as the number of trials generated by subject, by label. 'Idle' and 'Rest' trials were generated by binning across entire label durations, while 'Left', 'Right', 'Joint', and 'Move' trials were generated with respect to the movement onset.

Subject No.	Idle	Rest	Left	Right	Joint	Move
1	2303	138	67	51	17	118
2	9534	7670	227	278	87	505
3	17650	24777	216	374	75	590

Table 4.9 Sensitivity and specificity reports for each subject follow the format shown below. For each paradigm, class-specific sensitivity and specificity ($\mu \pm \sigma$) of each model are reported. Mean and standard deviation is calculated from 5-fold cross-validation. Examples of 2-class and 3-class paradigm reporting are shown below.

Subject N	LDA	SVM	MLP
Labels	Sensitivity/Specificity	Sensitivity/Specificity	Sensitivity/Specificity
*Class*1	$\mu_{sn} \pm \sigma_{sn}$ / $\mu_{sp} \pm \sigma_{sp}$	$\mu_{sn} \pm \sigma_{sn}$ / $\mu_{sp} \pm \sigma_{sp}$	$\mu_{sn} \pm \sigma_{sn}$ / $\mu_{sp} \pm \sigma_{sp}$
*Class*2	...	...	...
*Class*1	$\mu_{sn} \pm \sigma_{sn}$ / $\mu_{sp} \pm \sigma_{sp}$	$\mu_{sn} \pm \sigma_{sn}$ / $\mu_{sp} \pm \sigma_{sp}$	$\mu_{sn} \pm \sigma_{sn}$ / $\mu_{sp} \pm \sigma_{sp}$
*Class*2	...	...	...
*Class*3	...	...	...

Table 4.10 Sensitivity and specificity reports for Subject 1, for each paradigm and model. We report mean and standard deviation across 5 folds for each class.

Subject 1	LDA	SVM	MLP
Labels	Sensitivity/Specificity	Sensitivity/Specificity	Sensitivity/Specificity
Rest	0.66 ± 0.01 / 0.54 ± 0.05	0.01 ± 0.02 / 1.00 ± 0.00	0.38 ± 0.06 / 0.69 ± 0.07
Idle	0.54 ± 0.05 / 0.66 ± 0.10	1.00 ± 0.00 / 0.01 ± 0.02	0.69 ± 0.07 / 0.38 ± 0.06
Idle	0.79 ± 0.10 / 0.74 ± 0.07	0.4 ± 0.08 / 0.97 ± 0.03	0.75 ± 0.07 / 0.71 ± 0.06
$Move_{pre}$	0.35 ± 0.06 / 0.84 ± 0.08	0.04 ± 0.02 / 0.99 ± 0.01	0.27 ± 0.09 / 0.84 ± 0.01
$Move_{on}$	0.57 ± 0.07 / 0.92 ± 0.03	0.97 ± 0.03 / 0.37 ± 0.09	0.55 ± 0.10 / 0.89 ± 0.07
$Left_{all}$	0.60 ± 0.13 / 0.75 ± 0.13	0.63 ± 0.14 / 0.55 ± 0.15	0.67 ± 0.22 / 0.49 ± 0.13
$Right_{all}$	0.75 ± 0.13 / 0.60 ± 0.13	0.55 ± 0.15 / 0.63 ± 0.14	0.49 ± 0.13 / 0.67 ± 0.22
$Left_{pre}$	0.76 ± 0.10 / 0.76 ± 0.10	0.76 ± 0.11 / 0.51 ± 0.09	0.73 ± 0.14 / 0.45 ± 0.27
$Right_{pre}$	0.76 ± 0.10 / 0.76 ± 0.10	0.51 ± 0.09 / 0.76 ± 0.11	0.45 ± 0.27 / 0.73 ± 0.14
$Right_{pre}$	0.67 ± 0.15 / 0.69 ± 0.09	0.51 ± 0.19 / 0.56 ± 0.21	0.38 ± 0.38 / 0.76 ± 0.25
$Right_{on}$	0.69 ± 0.09 / 0.67 ± 0.15	0.56 ± 0.21 / 0.51 ± 0.19	0.76 ± 0.25 / 0.38 ± 0.38
$Left_{pre}$	0.82 ± 0.10 / 0.78 ± 0.08	0.70 ± 0.15 / 0.69 ± 0.03	0.57 ± 0.27 / 0.66 ± 0.13
$Left_{on}$	0.78 ± 0.08 / 0.82 ± 0.10	0.69 ± 0.03 / 0.70 ± 0.15	0.66 ± 0.13 / 0.57 ± 0.27
Rest	0.55 ± 0.10 / 0.66 ± 0.08	0.04 ± 0.04 / 0.96 ± 0.03	0.26 ± 0.21 / 0.79 ± 0.06
Idle	0.43 ± 0.12 / 0.80 ± 0.05	0.93 ± 0.04 / 0.30 ± 0.09	0.58 ± 0.07 / 0.63 ± 0.07
$Move_{pre}$	0.21 ± 0.03 / 0.87 ± 0.06	0.06 ± 0.06 / 0.98 ± 0.01	0.19 ± 0.10 / 0.87 ± 0.04
$Move_{on}$	0.64 ± 0.10 / 0.90 ± 0.06	0.41 ± 0.09 / 0.96 ± 0.03	0.53 ± 0.06 / 0.92 ± 0.04
Idle	0.68 ± 0.13 / 0.76 ± 0.08	0.99 ± 0.01 / 0.22 ± 0.07	0.71 ± 0.05 / 0.71 ± 0.09
$Left_{pre}$	0.37 ± 0.10 / 0.80 ± 0.06	0.01 ± 0.03 / 1.00 ± 0.00	0.15 ± 0.10 / 0.86 ± 0.02
$Left_{on}$	0.52 ± 0.04 / 0.96 ± 0.02	0.22 ± 0.08 / 0.99 ± 0.01	0.34 ± 0.08 / 0.95 ± 0.02
$Right_{pre}$	0.23 ± 0.11 / 0.96 ± 0.02	0.04 ± 0.05 / 1.00 ± 0.00	0.23 ± 0.10 / 0.93 ± 0.02
$Right_{on}$	0.28 ± 0.17 / 0.95 ± 0.05	0.04 ± 0.05 / 0.99 ± 0.02	0.25 ± 0.16 / 0.94 ± 0.02

Table 4.11 Sensitivity and specificity reports for Subject 2, for each paradigm and model. We report mean and standard deviation across 5 folds for each class.

Subject 2	LDA	SVM	MLP
Labels	Sensitivity/Specificity	Sensitivity/Specificity	Sensitivity/Specificity
Rest	0.59 ± 0.310 / 0.79 ± 0.199	0.48 ± 0.305 / 0.71 ± 0.148	0.56 ± 0.229 / 0.68 ± 0.187
Still	0.79 ± 0.199 / 0.59 ± 0.310	0.71 ± 0.148 / 0.48 ± 0.305	0.68 ± 0.187 / 0.56 ± 0.229
Idle	0.98 ± 0.014 / 0.58 ± 0.047	0.90 ± 0.063 / 0.74 ± 0.056	0.77 ± 0.080 / 0.75 ± 0.080
$Move_{pre}$	0.22 ± 0.040 / 0.97 ± 0.014	0.27 ± 0.041 / 0.92 ± 0.050	0.41 ± 0.054 / 0.82 ± 0.061
$Move_{on}$	0.45 ± 0.061 / 0.98 ± 0.004	0.58 ± 0.056 / 0.95 ± 0.020	0.43 ± 0.107 / 0.93 ± 0.028
$Left_{all}$	0.56 ± 0.063 / 0.66 ± 0.056	0.56 ± 0.120 / 0.61 ± 0.078	0.54 ± 0.116 / 0.59 ± 0.100
$Right_{all}$	0.66 ± 0.056 / 0.56 ± 0.063	0.61 ± 0.078 / 0.56 ± 0.120	0.59 ± 0.100 / 0.54 ± 0.116
$Left_{pre}$	0.58 ± 0.122 / 0.66 ± 0.126	0.55 ± 0.129 / 0.63 ± 0.128	0.47 ± 0.116 / 0.67 ± 0.090
$Right_{pre}$	0.66 ± 0.126 / 0.58 ± 0.122	0.63 ± 0.128 / 0.55 ± 0.129	0.67 ± 0.090 / 0.47 ± 0.116
$Right_{pre}$	0.79 ± 0.048 / 0.42 ± 0.052	0.60 ± 0.074 / 0.52 ± 0.122	0.51 ± 0.084 / 0.61 ± 0.054
$Right_{on}$	0.42 ± 0.052 / 0.79 ± 0.048	0.52 ± 0.122 / 0.60 ± 0.074	0.61 ± 0.054 / 0.51 ± 0.084
$Left_{pre}$	0.77 ± 0.079 / 0.48 ± 0.070	0.66 ± 0.076 / 0.50 ± 0.065	0.52 ± 0.078 / 0.63 ± 0.090
$Left_{on}$	0.48 ± 0.070 / 0.77 ± 0.079	0.50 ± 0.065 / 0.66 ± 0.076	0.63 ± 0.090 / 0.52 ± 0.078
Rest	0.57 ± 0.326 / 0.80 ± 0.169	0.47 ± 0.306 / 0.79 ± 0.173	0.55 ± 0.213 / 0.72 ± 0.116
Idle	0.77 ± 0.180 / 0.58 ± 0.282	0.68 ± 0.180 / 0.53 ± 0.253	0.48 ± 0.138 / 0.69 ± 0.139
$Move_{pre}$	0.20 ± 0.029 / 0.98 ± 0.006	0.25 ± 0.025 / 0.95 ± 0.029	0.37 ± 0.080 / 0.86 ± 0.072
$Move_{on}$	0.45 ± 0.066 / 0.99 ± 0.003	0.61 ± 0.037 / 0.97 ± 0.011	0.49 ± 0.096 / 0.95 ± 0.024
Idle	0.96 ± 0.020 / 0.59 ± 0.031	0.89 ± 0.070 / 0.72 ± 0.067	0.63 ± 0.075 / 0.81 ± 0.053
$Left_{pre}$	0.13 ± 0.047 / 0.99 ± 0.004	0.15 ± 0.019 / 0.97 ± 0.018	0.23 ± 0.062 / 0.91 ± 0.030
$Left_{on}$	0.41 ± 0.028 / 0.98 ± 0.001	0.40 ± 0.053 / 0.97 ± 0.007	0.30 ± 0.091 / 0.95 ± 0.009
$Right_{pre}$	0.13 ± 0.041 / 0.96 ± 0.018	0.14 ± 0.046 / 0.94 ± 0.042	0.27 ± 0.043 / 0.84 ± 0.046
$Right_{on}$	0.11 ± 0.042 / 0.99 ± 0.003	0.20 ± 0.057 / 0.97 ± 0.012	0.23 ± 0.059 / 0.90 ± 0.028

Table 4.12 Sensitivity and specificity reports for Subject 3, for each paradigm and model. We report mean and standard deviation across 5 folds for each class.

Subject 3	LDA	SVM	MLP
Labels	Sensitivity/Specificity	Sensitivity/Specificity	Sensitivity/Specificity
Rest	0.58 ± 0.187 / 0.57 ± 0.165	0.61 ± 0.175 / 0.59 ± 0.105	0.51 ± 0.233 / 0.63 ± 0.216
Still	0.57 ± 0.165 / 0.58 ± 0.187	0.59 ± 0.105 / 0.61 ± 0.175	0.63 ± 0.216 / 0.51 ± 0.233
Idle	0.89 ± 0.053 / 0.40 ± 0.113	0.90 ± 0.021 / 0.53 ± 0.113	0.65 ± 0.057 / 0.71 ± 0.061
$Move_{pre}$	0.28 ± 0.055 / 0.90 ± 0.050	0.30 ± 0.085 / 0.94 ± 0.009	0.35 ± 0.115 / 0.82 ± 0.040
$Move_{on}$	0.23 ± 0.099 / 0.98 ± 0.012	0.25 ± 0.095 / 0.94 ± 0.015	0.42 ± 0.060 / 0.81 ± 0.050
$Left_{all}$	0.54 ± 0.177 / 0.57 ± 0.166	0.33 ± 0.096 / 0.67 ± 0.043	0.42 ± 0.138 / 0.61 ± 0.100
$Right_{all}$	0.57 ± 0.166 / 0.54 ± 0.177	0.67 ± 0.043 / 0.33 ± 0.096	0.61 ± 0.100 / 0.42 ± 0.138
$Left_{pre}$	0.49 ± 0.166 / 0.61 ± 0.149	0.32 ± 0.077 / 0.70 ± 0.063	0.36 ± 0.134 / 0.69 ± 0.106
$Right_{pre}$	0.61 ± 0.149 / 0.49 ± 0.166	0.70 ± 0.063 / 0.32 ± 0.077	0.69 ± 0.106 / 0.36 ± 0.134
$Right_{pre}$	0.46 ± 0.110 / 0.72 ± 0.075	0.52 ± 0.120 / 0.54 ± 0.030	0.56 ± 0.239 / 0.52 ± 0.302
$Right_{on}$	0.72 ± 0.075 / 0.46 ± 0.110	0.54 ± 0.030 / 0.52 ± 0.120	0.52 ± 0.302 / 0.56 ± 0.239
$Left_{pre}$	0.72 ± 0.103 / 0.54 ± 0.055	0.63 ± 0.122 / 0.54 ± 0.163	0.58 ± 0.236 / 0.55 ± 0.185
$Left_{on}$	0.54 ± 0.055 / 0.72 ± 0.103	0.54 ± 0.163 / 0.63 ± 0.122	0.55 ± 0.185 / 0.58 ± 0.236
Rest	0.60 ± 0.183 / 0.52 ± 0.143	0.57 ± 0.193 / 0.65 ± 0.105	0.41 ± 0.144 / 0.73 ± 0.108
Idle	0.45 ± 0.160 / 0.64 ± 0.193	0.57 ± 0.107 / 0.59 ± 0.178	0.44 ± 0.052 / 0.60 ± 0.095
$Move_{pre}$	0.21 ± 0.072 / 0.96 ± 0.024	0.27 ± 0.092 / 0.97 ± 0.006	0.38 ± 0.062 / 0.86 ± 0.041
$Move_{on}$	0.20 ± 0.076 / 0.99 ± 0.006	0.18 ± 0.089 / 0.98 ± 0.006	0.38 ± 0.108 / 0.92 ± 0.028
Idle	0.50 ± 0.208 / 0.78 ± 0.102	0.90 ± 0.029 / 0.47 ± 0.104	0.56 ± 0.087 / 0.79 ± 0.050
$Left_{pre}$	0.16 ± 0.106 / 0.94 ± 0.040	0.06 ± 0.048 / 0.98 ± 0.003	0.13 ± 0.104 / 0.90 ± 0.031
$Left_{on}$	0.36 ± 0.145 / 0.83 ± 0.078	0.07 ± 0.037 / 0.99 ± 0.004	0.22 ± 0.101 / 0.92 ± 0.030
$Right_{pre}$	0.18 ± 0.102 / 0.87 ± 0.049	0.15 ± 0.053 / 0.95 ± 0.012	0.24 ± 0.062 / 0.84 ± 0.044
$Right_{on}$	0.26 ± 0.197 / 0.86 ± 0.142	0.12 ± 0.031 / 0.96 ± 0.015	0.24 ± 0.105 / 0.88 ± 0.047

Table 4.13 All levels of baseline chance used to evaluate each model. Estimating finite chance levels uses number of trials, number of classes, and level of statistical significance as parameters, and are also included below.

Subject No.	Classification Paradigm	N_{trials}	$N_{classes}$	inf	fin. ($p \leq 0.05$)	fin. ($p \leq 0.01$)
1	Still-Rest	2441	2	0.50	0.52	0.52
	Still-Move-pre-Move-on	2573	3	0.33	0.35	0.36
	Left-Right	118	2	0.50	0.58	0.61
	Left-Left-on	134	2	0.50	0.57	0.60
	Right-Right-on	102	2	0.50	0.58	0.62
	Left-all-Right-all	236	2	0.50	0.56	0.58
	Rest-Still-Move-pre-Move-on	2711	4	0.25	0.26	0.27
	Still-Left-Left-on-Right-Right-on	2539	5	0.20	0.21	0.22
2	Still-Rest	17204	2	0.50	0.51	0.51
	Still-Move-pre-Move-on	10720	3	0.33	0.34	0.34
	Left-Right	505	2	0.50	0.54	0.55
	Left-Left-on	455	2	0.50	0.54	0.55
	Right-Right-on	557	2	0.50	0.54	0.55
	Left-all-Right-all	1012	2	0.50	0.53	0.54
	Rest-Still-Move-pre-Move-on	18390	4	0.25	0.26	0.26
	Still-Left-Left-on-Right-Right-on	10546	5	0.20	0.21	0.21
3	Still-Rest	42427	2	0.50	0.50	0.51
	Still-Move-pre-Move-on	18980	3	0.33	0.34	0.34
	Left-Right	590	2	0.50	0.53	0.55
	Left-Left-on	432	2	0.50	0.54	0.56
	Right-Right-on	748	2	0.50	0.53	0.54
	Left-all-Right-all	1180	2	0.50	0.52	0.53
	Rest-Still-Move-pre-Move-on	43757	4	0.25	0.25	0.25
	Still-Left-Left-on-Right-Right-on	18830	5	0.20	0.20	0.21

4.8 Supplementary Figures

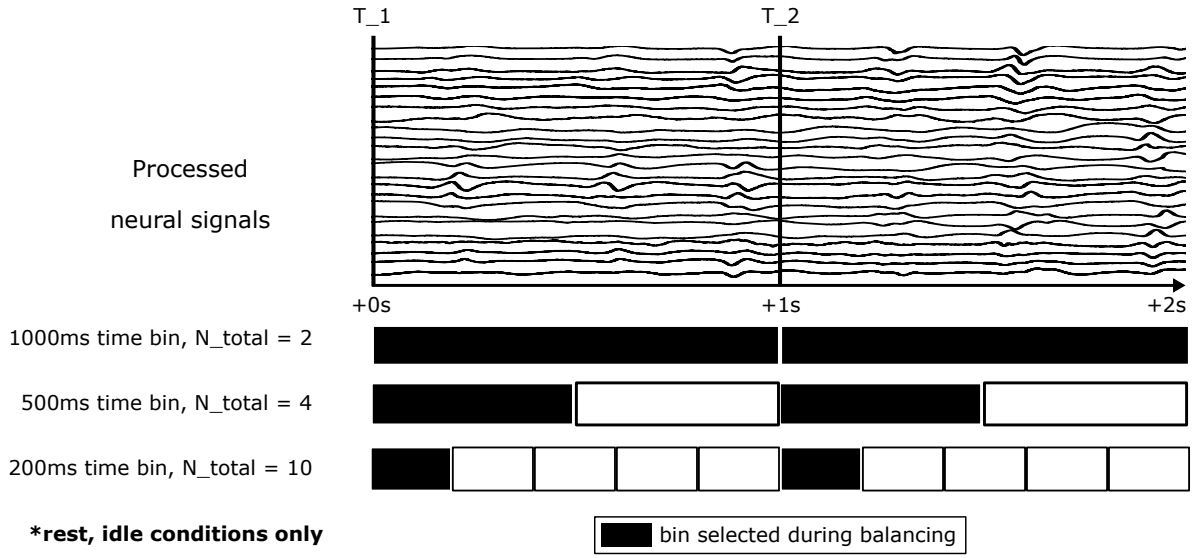


Figure 4.10 **Balancing baseline features across window sizes.** Sampling the same baseline period with decreasing window sizes yields an increasing number of samples for that same period. To account for this imbalance, baseline features at every window size were down-sampled to match the number of features generated from the largest (1000*ms*) window size. Specifically, the features at each window size were matched by nearest time stamp to each baseline feature generated with the largest (1000*ms*) window. This ensured that baseline features at each window size represented similar points in time relative to the end of their respective sampling windows.

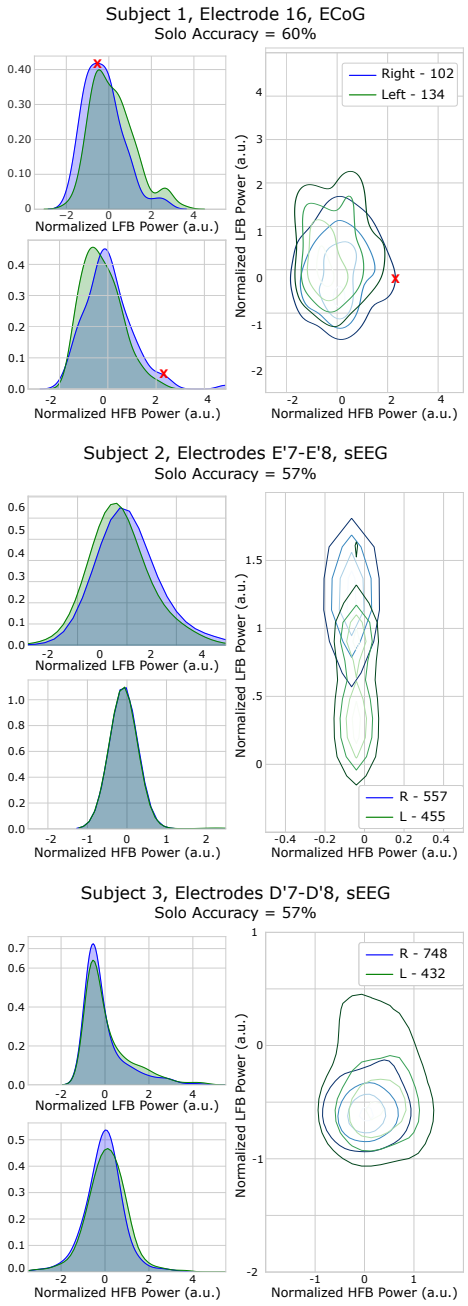


Figure 4.11 **Example of neural features from top electrode over sensorimotor region (all Subjects).** Distributions of 8-32Hz low frequency band (LFB) and 76-100Hz high frequency band (HFB) RMS-power. The total number of epochs for each class are reported in the legend. Subject 1 was sampled from a window size of 200*ms*, and taken 350*ms* prior to movement onset. Subject 2 was sampled from a window size of 1000*ms*, and taken 250*ms* prior to movement onset. Subject 3 was sampled from a window size of 800*ms*, and taken 250*ms* prior to movement onset.

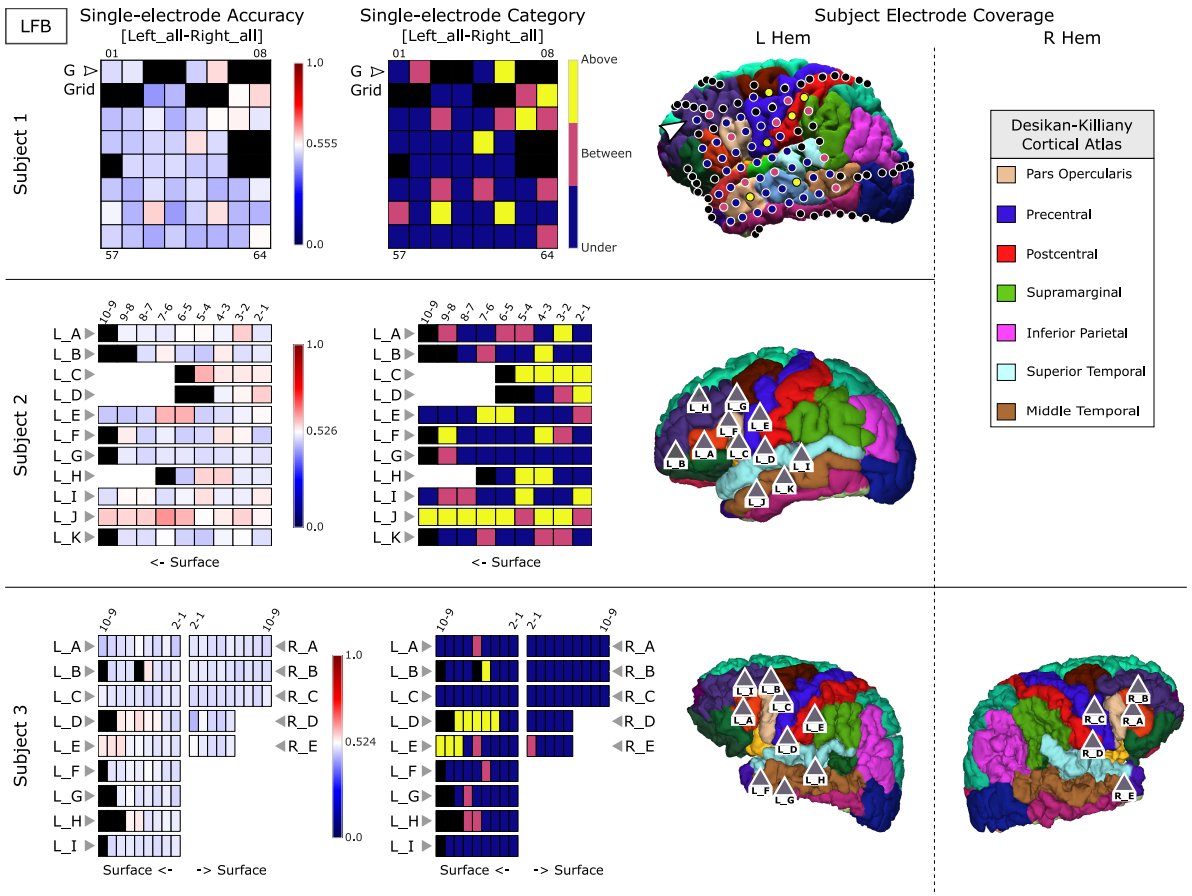


Figure 4.12 **Mapping single-electrode scores for 'Left$_{all}$' vs 'Right$_{all}$' onset classification, using only LFB.** Each electrode was scored by the ability of their LFB features to correctly classify movement onset labels, such as 'Left$_{all}$' vs 'Right$_{all}$'. The first column shows each electrode score averaged across 5-folds cross-validation, showing which electrodes scored above (red) the less stringent 'infinite' chance level using a divergent color map. Electrode references are also provided to orient onto each subject's electrode reconstruction map (third, fourth column). In the second column, electrodes were grouped into three discrete categories based on how their score compared to both 'infinite' and 'finite' levels of chance. (pink) Electrodes with a solo classification score above the less stringent 'infinite'-level of chance, but below the more stringent 'finite'-level of chance, suggest marginal distinction between features; (yellow) electrodes with a solo classification score above the 'finite'-level of chance suggest linearly distinguishable distinctions in estimated band power between movement labels; (blue) electrodes with solo classification below both baselines suggest no distinction between labels, 'Left$_{all}$' vs 'Right$_{all}$'. For all subjects, these electrodes were identified over or near central sulcus. The third and fourth columns show the post-implant electrode reconstruction maps for each subject. Subjects with ECoG coverage (Subject 1) show electrode responses mapped to the cortical surface. Subjects with sEEG coverage (Subjects 2,3) show where along the cortical surface each probe was inserted.

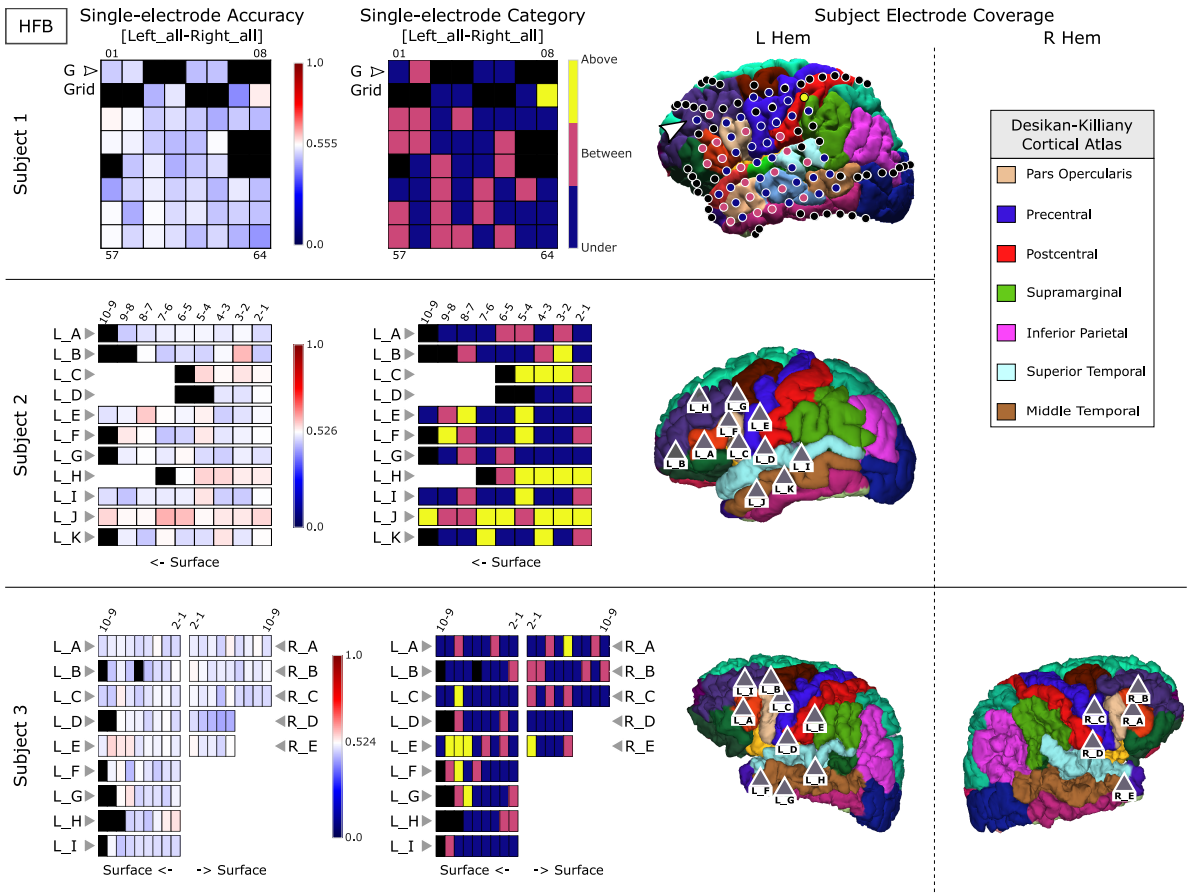


Figure 4.13 **Mapping single-electrode scores for 'Left$_{all}$' vs 'Right$_{all}$' onset classification, using only HFB.** Each electrode was scored by the ability of their HFB features to correctly classify movement onset labels, such as 'Left$_{all}$' vs 'Right$_{all}$'. The first column shows each electrode score averaged across 5-folds cross-validation, showing which electrodes scored above (red) the less stringent 'infinite' chance level using a divergent color map. Electrode references are also provided to orient onto each subject's electrode reconstruction map (third, fourth column). In the second column, electrodes were grouped into three discrete categories based on how their score compared to both 'infinite' and 'finite' levels of chance. (pink) Electrodes with a solo classification score above the less stringent 'infinite'-level of chance, but below the more stringent 'finite'-level of chance, suggest marginal distinction between features; (yellow) electrodes with a solo classification score above the 'finite'-level of chance suggest linearly distinguishable distinctions in estimated band power between movement labels; (blue) electrodes with solo classification below both baselines suggest no distinction between labels, 'Left$_{all}$' vs 'Right$_{all}$'. For all subjects, these electrodes were identified over or near central sulcus. The third and fourth columns show the post-implant electrode reconstruction maps for each subject. Subjects with ECoG coverage (Subject 1) show electrode responses mapped to the cortical surface. Subjects with sEEG coverage (Subjects 2,3) show where along the cortical surface each probe was inserted.

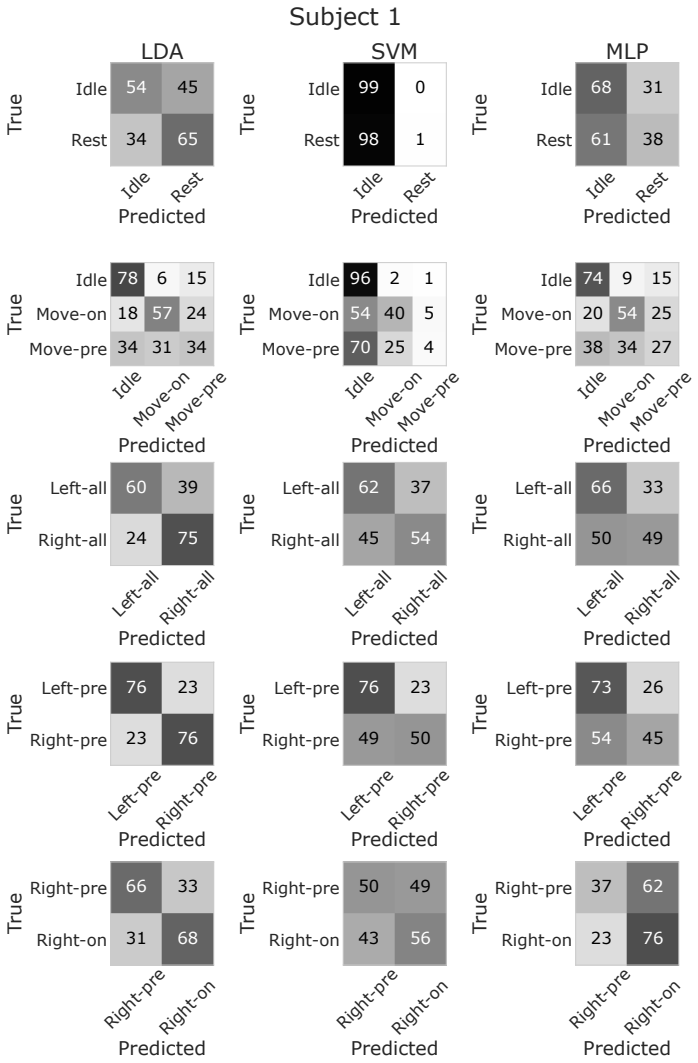


Figure 4.14 **Best model scores for each classifier paradigm, Subject 1.** Linear Discriminant Analysis (LDA), Support Vector Machine (SVM), and Multilayer Perceptron (MLP) classifiers were trained on features from movement ('Left', 'Right', or 'Move') and non-movement periods ('Idle', 'Rest'). Uniform class priors were enforced while training each model, and test scores were unweighted to allow comparison between classifier paradigms. Classifier accuracies are reported as confusion matrices that summarize model evaluation over 5-fold cross validation. Confusion matrices are normalized so each value represents N% classification performance, rounded to the nearest digit.

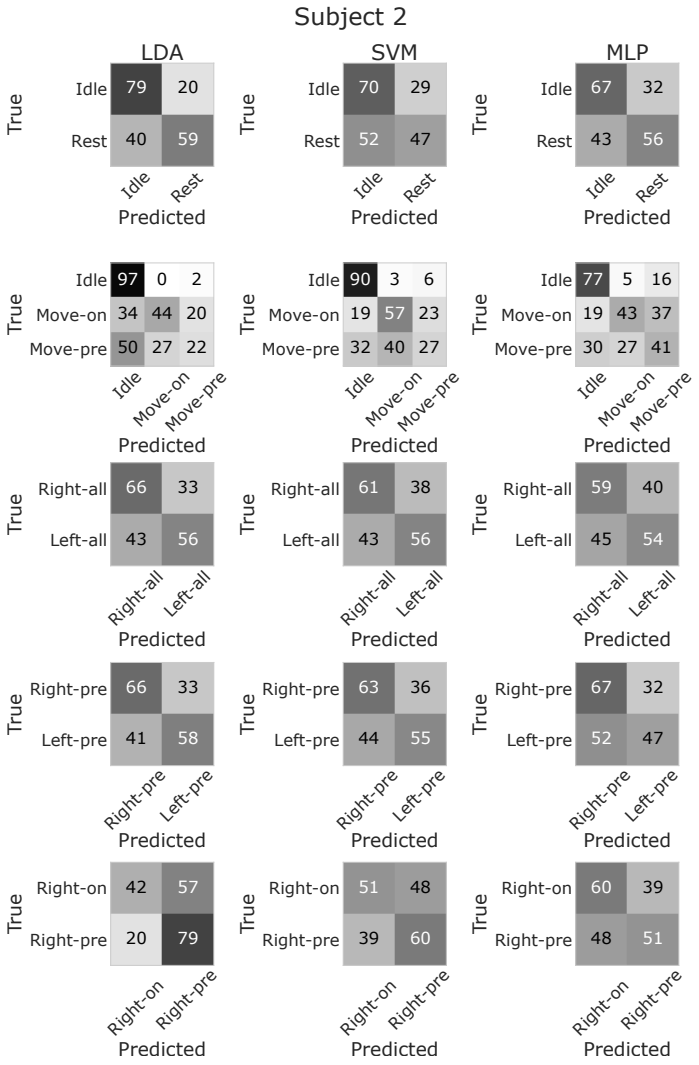


Figure 4.15 **Best model scores for each classifier paradigm, Subject 2.** Description available in Figure 4.14.

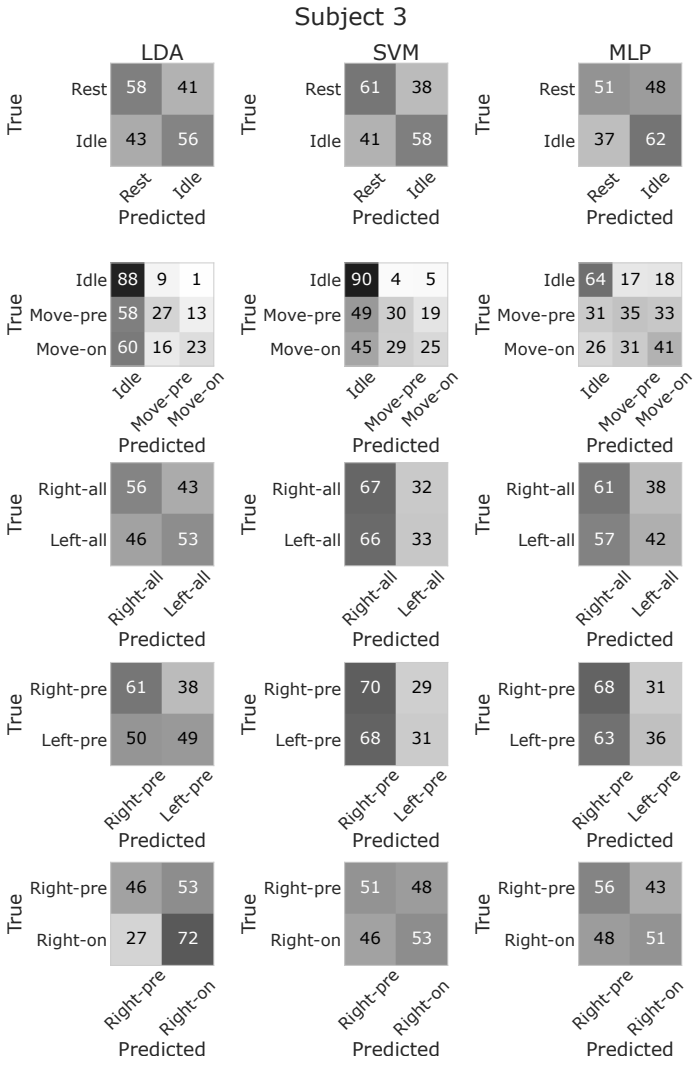


Figure 4.16 **Best model scores for each classifier paradigm, Subject 3.** Description available in Figure 4.14.

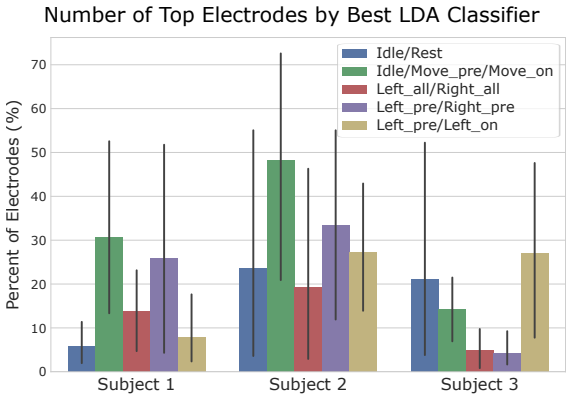


Figure 4.17 **Percent of electrodes used by best LDA classifiers.** A key hyperparameter for LDA classifier training and evaluation was the number of input features for each classifier variant. HFB and LFB features were selected sequentially, starting from the electrode with highest solo accuracy for a given classification paradigm. The mean and variance of the number of electrodes selected by the best LDA classifiers is reported by the bar plot with error lines, for each subject and classification paradigm.

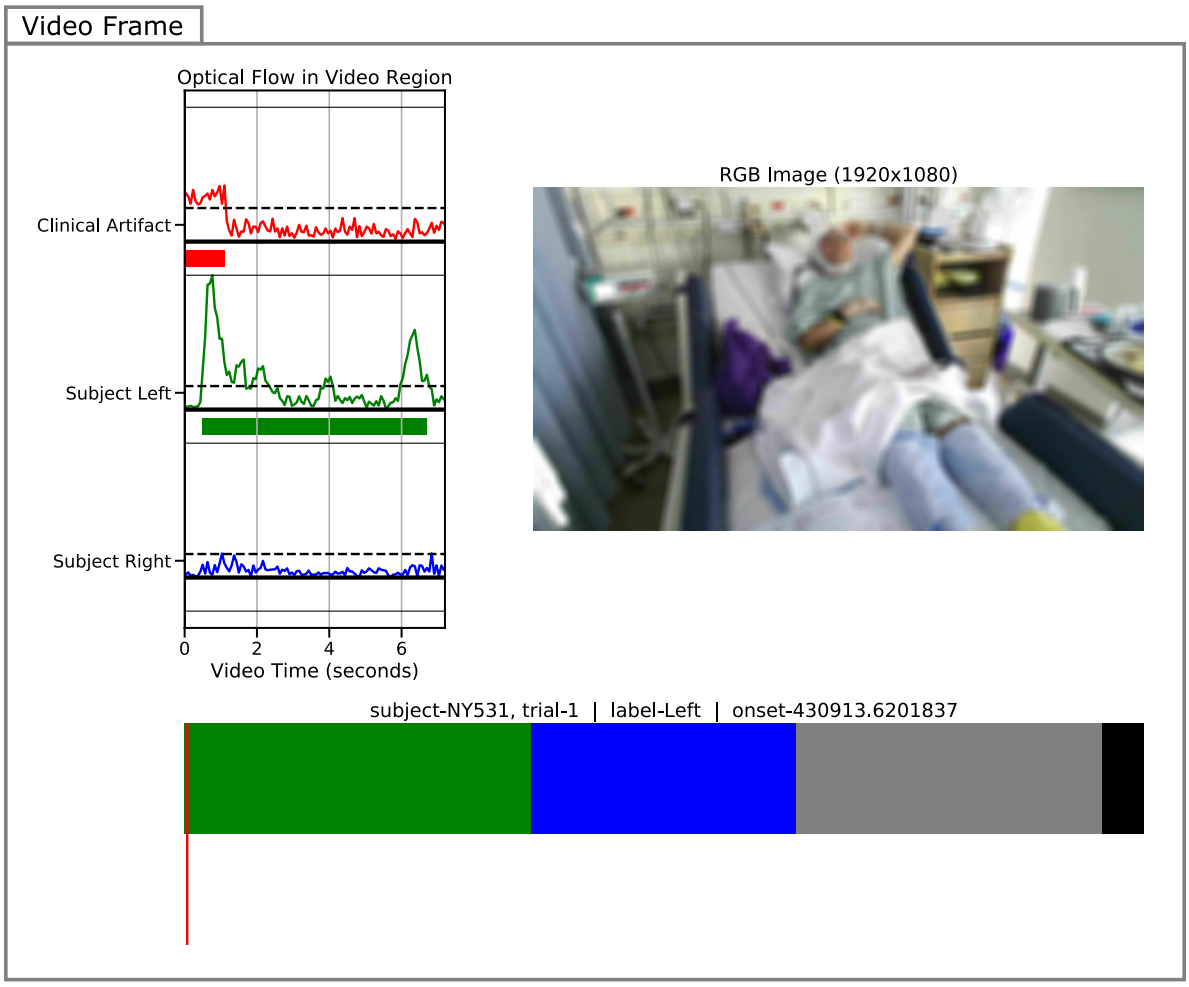


Figure 4.18 **Videos of subject behaviors show unstructured, uninstructed movements**. Example frame of review videos provided for Subjects 1, 2, and 3 (supp_V1.mp4, supp_V2.mp4, and supp_V3.mp4). Each video consists of 1-3 examples each of Left (green), Right (blue), Idle/Still (gray), and Rest (black) periods identified by our semi-automated methods. A variety of behaviors are shown, including basic reaching and scratching movements. Examples where clinical care is being provided are also shown. Finally, a few limits of the coarse labeling process are demonstrated for Subject 3 (RCH3). *Top left*: Motion report for a given trial, showing magnitude of flow consolidated in Left, Right, and Clinical Care (red) ROIs. Beneath each flow signal is the movement segment as labeled by our methods. The vertical black line traces video playback. *Top right*: Video segment for a given trial as recorded by the Kinect v2 color (RGB) camera. Recorded images are blurred in post to protect patient confidentiality. *Bottom*: Legend describes which subject, behavior label, and recording time the given trial corresponds to. For visualization purposes, trials were arranged by label and then onset time, and the red vertical line marks the given trial.

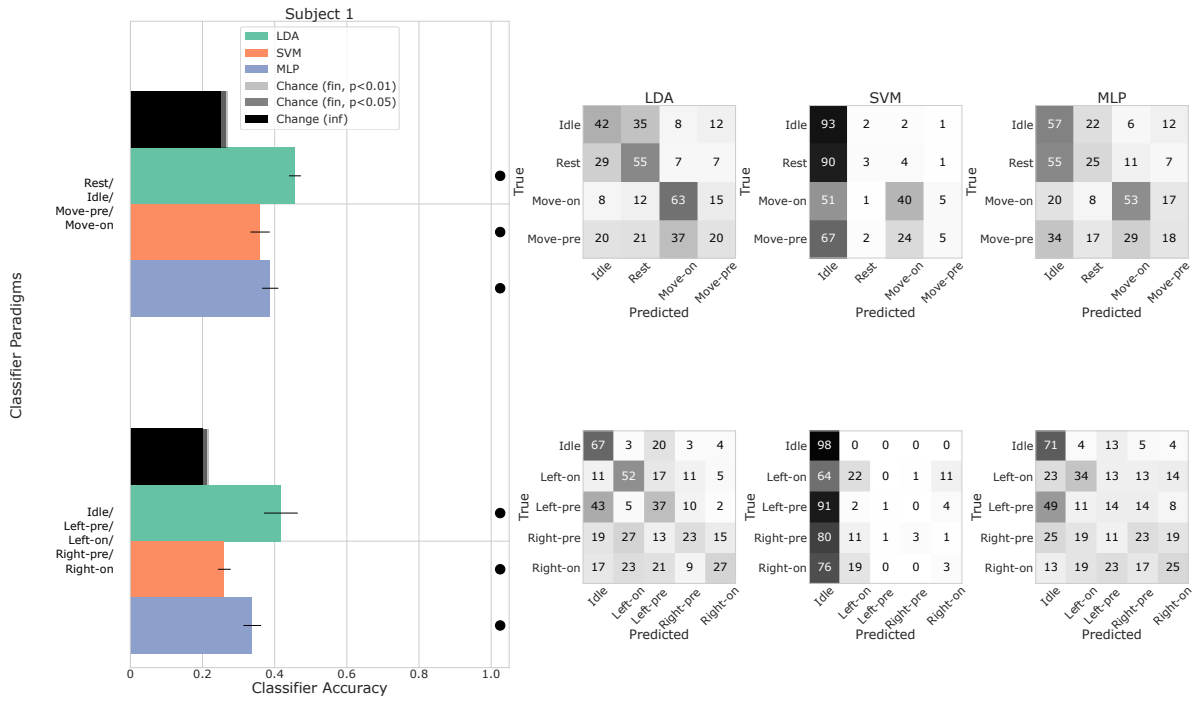


Figure 4.19 **Best model scores for the four- and five-class paradigms, Subject 1.** Linear Discriminant Analysis (LDA), Support Vector Machine (SVM), and Multilayer Perceptron (MLP) classifiers were evaluated on larger subsets of movement ('Left', 'Right', or 'Move') and non-movement ('Idle', 'Rest') labels. Uniform class priors were enforced while training each model, and test sets were unweighted to allow comparison between classifier paradigms. Classifier accuracies are reported as confusion matrices that summarize model evaluation over 5-fold cross validation. Confusion matrices are normalized so each value represents N% classification performance, rounded to the nearest digit.

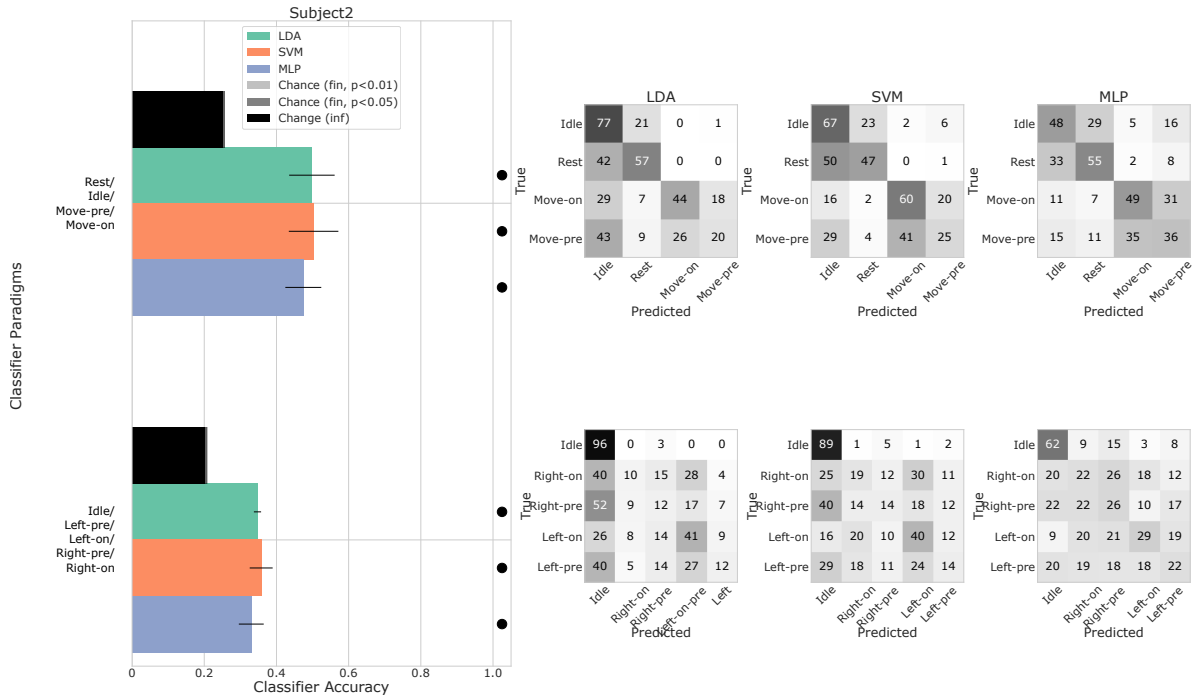


Figure 4.20 **Best model scores for the four- and five-class paradigms, Subject 2.** Description available in Figure 4.19.

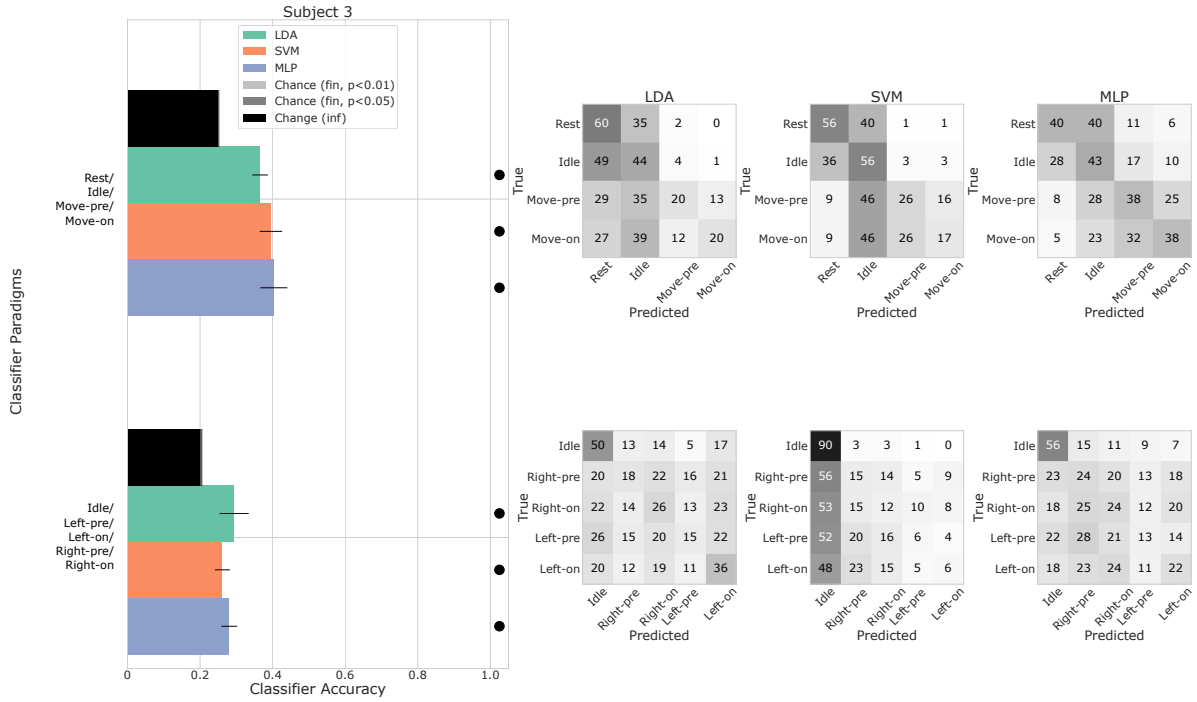


Figure 4.21 **Best model scores for the four- and five-class paradigms, Subject 3.** Description available in Figure 4.19.

References

[1] Anna Buttfield, Pierre W Ferrez, and Jd R Millan. Towards a robust bci: error potentials and online learning. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 14(2):164–168, 2006.

[2] Leigh R Hochberg, Daniel Bacher, Beata Jarosiewicz, Nicolas Y Masse, John D Simeral, Joern Vogel, Sami Haddadin, Jie Liu, Sydney S Cash, Patrick van der Smagt, et al. Reach and grasp by people with tetraplegia using a neurally controlled robotic arm. *Nature*, 485(7398):372, 2012.

[3] Inchul Choi, Ilsun Rhiu, Yushin Lee, Myung Hwan Yun, and Chang S Nam. A systematic review of hybrid brain-computer interfaces: Taxonomy and usability perspectives. *PloS one*, 12(4):e0176674, 2017.

[4] Gerwin Schalk, Eric C Leuthardt, Peter Brunner, Jeffrey G Ojemann, Lester A Gerhardt, and Jonathan R Wolpaw. Real-time detection of event-related brain activity. *Neuroimage*, 43(2):245–249, 2008.

[5] Korostenskaja M, Wilson A, Rose D, Brunner P, Schalk G, Leach J, Mangano F, Fujiwara H, Rozhkov L, Harris E, et al. Real-time functional mapping with electrocorticography in pediatric epilepsy: comparison with fmri and esm findings. *Clinical EEG and Neuroscience*, **45**(3):205–211, 2014.

[6] Peter Brunner, Anthony L Ritaccio, Timothy M Lynch, Joseph F Emrich, J Adam Wilson, Justin C Williams, Erik J Aarnoutse, Nick F Ramsey, Eric C Leuthardt, Horst Bischof, et al. A practical procedure for real-time functional mapping of eloquent cortex using electrocorticographic signals in humans. *Epilepsy & Behavior*, 15(3):278–286, 2009.

[7] Mariska J Vansteensel, Elmar GM Pels, Martin G Bleichner, Mariana P Branco, Timothy Denison, Zachary V Freudenburg, Peter Gosselaar, Sacha Leinders, Thomas H Ottens, Max A Van Den Boom, et al. Fully implanted brain–computer interface in a locked-in patient with als. *New England Journal of Medicine*, 375(21):2060–2066, 2016.

[8] Fetz E. Operant conditioning of cortical unit activity. *Science*, **163**(3870):955–958, 1969.

[9] Bouchard K, Mesgarani N, Johnson K, and Chang E. Functional organization of human sensorimotor cortex for speech articulation. *Nature*, **495**(7441):327–332, March 2013.

[10] Mesgarani N and Chang E. Selective cortical representation of attended speaker in multi-talker speech perception. *Nature*, **485**(7397):233–6, May 2012.

[11] Miguel Angrick, Christian Herff, Emily Mugler, Matthew C Tate, Marc W Slutzky, Dean J Krusienski, and Tanja Schultz. Speech synthesis from ecog using densely connected 3d convolutional neural networks. *bioRxiv*, page 478644, 2018.

[12] Wang W, Collinger J, Degenhart A, Tyler-Kabara E, Schwartz A, Moran D, Weber D, Wodlinger B, Vinjamuri R, Ashmore R, Kelly J, and Boninger M. An Electrocorticographic Brain Interface in an Individual with Tetraplegia. *PLoS ONE*, **8**, 2013.

[13] Carmena J, Lebedev M, Crist R, O'Doherty J, Santucci D, Dimitrov D, Patil P, Henriquez C, and Nicolelis M. Learning to control a brain–machine interface for reaching and grasping by primates. *PLoS Biology*, **1**(2), 10 2003.

[14] Chaplin J, Moxon K, Markowitz R, and Nicolelis M. Real-time control of a robot arm using simultaneously recorded neurons in the motor cortex. *Nature*, **2**:664–670, 1999.

[15] Gilja V, Nuyujukian P, Chestek C, Cunningham J, Krishna S, et al. A high-performance neural prosthesis enabled by control algorithm design. *Nature Neuroscience*, **15**:1752–1758, 2012.

[16] Hochberg L, Serruya M, Friehs G, Mukand J, Saleh M, Caplan A, Branner A, Chen D, Penn R, and Donoghue J. Neuronal ensemble control of prosthetic devices by a human with tetraplegia. *Nature*, **442**:164–171, 2006.

[17] Velliste M, Perel S, Spalding C, Whitford A, and Schwartz A. Cortical control of a prosthetic arm for self-feeding. *Nature*, **458**:1098–1101, 2008.

[18] György Buzsáki, Costas A Anastassiou, and Christof Koch. The origin of extracellular fields and currents—eeg, ecog, lfp and spikes. *Nature reviews neuroscience*, 13(6):407, 2012.

[19] Scott R Cole and Bradley Voytek. Brain oscillations and the importance of waveform shape. *Trends in Cognitive Sciences*, 21(2):137–149, 2017.

[20] K Hara, S Uematsu, R Lesser, B Gordon, J Hart, and E Vining. Representation of primary motor cortex in humans: studied with chronic subdural grid. *Epilepsia*, 32(SUPPL.):23–24, 1991.

[21] Robert Prueckl, Christoph Kapeller, Cristhian Potes, Milena Korostenskaja, Gerwin Schalk, Ki H Lee, and Christoph Guger. cortiq-clinical software for electrocorticographic real-time functional mapping of the eloquent cortex. In *Conf Proc IEEE Eng Med Biol Soc*, volume 2013, pages 6365–6368, 2013.

[22] Kai J Miller, Eric C Leuthardt, Gerwin Schalk, Rajesh PN Rao, Nicholas R Anderson, Daniel W Moran, John W Miller, and Jeffrey G Ojemann. Spectral changes in cortical surface potentials during motor movement. *Journal of Neuroscience*, 27(9):2424–2432, 2007.

[23] Kai J Miller, Christopher J Honey, Dora Hermes, Rajesh PN Rao, Jeffrey G Ojemann, et al. Broadband changes in the cortical surface potential track activation of functionally diverse neuronal populations. *Neuroimage*, 85:711–720, 2014.

[24] Spencer Kellis, Kai Miller, Kyle Thomson, Richard Brown, Paul House, and Bradley Greger. Decoding spoken words using local field potentials recorded from the cortical surface. *Journal of neural engineering*, 7(5):056007, 2010.

[25] Dastjerdi M, Ozker M, Foster B, Rangarajan V, and Parvizi J. Numerical processing in the human parietal cortex during experimental and natural conditions. *Nature Communications*, **4**:2528, January 2013.

[26] Scott B Wilson, Christine A Turner, Ronald G Emerson, and Mark L Scheuer. Spike detection ii: automatic, perception-based detection and clustering. *Clinical neurophysiology*, 110(3):404–411, 1999.

[27] Mark L Scheuer and Scott B Wilson. Data analysis for continuous eeg monitoring in the icu: seeing the forest and the trees. *Journal of clinical neurophysiology*, 21(5):353–378, 2004.

[28] F Fürbass, P Ossenblok, M Hartmann, H Perko, AM Skupch, G Lindinger, L Elezi, E Pataraia, AJ Colon, C Baumgartner, et al. Prospective multi-center study of an automatic online seizure detection system for epilepsy monitoring units. *Clinical Neurophysiology*, 126(6):1124–1131, 2015.

[29] Christoph Kapeller, Milena Korostenskaja, Robert Prueckl, Po-Ching Chen, Ki Heyeong Lee, Michael Westerveld, Christine M Salinas, Jane C Cook, James E Baumgartner, and Christoph Guger. Cortiq-based real-time functional mapping for epilepsy surgery. *Journal of clinical neurophysiology*, 32(3):e12–e22, 2015.

[30] Gabriel P, Doyle W, Devinsky O, Friedman D, Thesen T, and Gilja V. Neural correlates to automatic behavior estimations from rgb-d video in epilepsy unit. In *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the*, pages 3402–3405. IEEE, 2016.

[31] Nancy XR Wang, Ali Farhadi, Rajesh PN Rao, and Bingni W Brunton. Ajile movement prediction: Multimodal deep learning for natural human neural recordings and video. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[32] Wang N, Olson J, Ojemann J, Rao R, and Brunton B. Unsupervised decoding of long-term, naturalistic human neural recordings with automated video and audio annotations. *Frontiers in human neuroscience*, **10**, 2016.

[33] Abdulwahab Alasfour, Paolo Gutierrez Gabriel, Xi Jiang, Isaac Shamie, Lucia Melloni, Thomas Thesen, Patricia Dugan, Daniel Friedman, Werner Doyle, Orrin Devinsky, et al. Coarse behavioral context decoding. *Journal of Neural Engineering*, 2018.

[34] Jamie Shotton, Toby Sharp, Alex Kipman, Andrew Fitzgibbon, Mark Finocchio, Andrew Blake, Mat Cook, and Richard Moore. Real-time human pose recognition in parts from single depth images. *Communications of the ACM*, 56(1):116–124, 2013.

[35] T. Pfister. *Advancing Human Pose and Gesture Recognition*. PhD thesis, University of Oxford, 2015.

[36] Rahul S Desikan, Florent Ségonne, Bruce Fischl, Brian T Quinn, Bradford C Dickerson, Deborah Blacker, Randy L Buckner, Anders M Dale, R Paul Maguire, Bradley T Hyman, et al. An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest. *Neuroimage*, 31(3):968–980, 2006.

[37] Quinn B, Carlson C, Doyle W, Cash S, Devinsky O, Spence C, Halgren E, and Thesen T. Intracranial cortical responses during visual-tactile integration in humans. *The Journal of Neuroscience : the official journal of the Society for Neuroscience*, **34**(1):171–181, January 2014.

[38] Gonzalez-Martinez J, Bulacio J, Thompson S, et al. Robotic epilepsy surgery: technique, results and complications related to robotic assisted seeg. *Neurosurgery*, **78**(2):169–179, 2016.

[39] Andrew R Dykstra, Alexander M Chan, Brian T Quinn, Rodrigo Zepeda, Corey J Keller, Justine Cormier, Joseph R Madsen, Emad N Eskandar, and Sydney S Cash. Individualized localization and cortical surface-based registration of intracranial electrodes. *Neuroimage*, 59(4):3563–3570, 2012.

[40] David Groppe, Stephan Bickel, Andrew Dykstra, Xiuyuan Wang, Pierre Mégevand, Manuel R. Mercier, Fred Lado, Ashesh Mehta, and Christopher Honey. ielvis: An open source matlab toolbox for localizing and visualizing human intracranial electrode data. 281, 02 2017.

[41] Gunnar Farnebäck. Two-frame motion estimation based on polynomial expansion. *Image analysis*, pages 363–370, 2003.

[42] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 3354–3361. IEEE, 2012.

[43] Y Liu, WG Coon, A De Pesters, P Brunner, and G Schalk. The effects of spatial filtering and artifacts on electrocorticographic signals. *Journal of neural engineering*, 12(5):056008, 2015.

[44] Gerwin Schalk, Joshua Marple, Robert T Knight, and William G Coon. Instantaneous voltage as an alternative to power-and phase-based interpretation of oscillatory brain activity. *Neuroimage*, 157:545–554, 2017.

[45] Katsuhiro Kobayashi, Julia Jacobs, and Jean Gotman. Detection of changes of high-frequency activity by statistical time-frequency analysis in epileptic spikes. *Clinical Neurophysiology*, 120(6):1070–1077, 2009.

[46] Juan R Vidal, Samson Freyermuth, Karim Jerbi, Carlos M Hamamé, Tomas Ossandon, Olivier Bertrand, Lorella Minotti, Philippe Kahane, Alain Berthoz, and Jean-Philippe Lachaux. Long-distance amplitude correlations in the high gamma band reveal segregation and integration within the reading network. *Journal of Neuroscience*, 32(19):6421–6434, 2012.

[47] Guangye Li, Shize Jiang, Sivylla E Paraskevopoulou, Meng Wang, Yang Xu, Zehan Wu, Liang Chen, Dingguo Zhang, and Gerwin Schalk. Optimal referencing for stereo-electroencephalographic (seeg) recordings. *NeuroImage*, 183:327–335, 2018.

[48] S Cole, T Donoghue, R Gao, and B Voytek. Neurodsp: A package for neural digital signal processing. *Journal of Open Source Software*, 4(36), 2019.

[49] Wei Wang, Alan D Degenhart, Jennifer L Collinger, Ramana Vinjamuri, Gustavo P Sudre, P David Adelson, Deborah L Holder, Eric C Leuthardt, Daniel W Moran, Michael L Boninger, et al. Human motor cortical activity recorded with micro-ecog electrodes, during individual finger movements. In *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 586–589. IEEE, 2009.

[50] Hai Sun, Timothy M Blakely, Felix Darvas, Jeremiah D Wander, Lise A Johnson, David K Su, Kai J Miller, Eberhard E Fetz, and Jeffery G Ojemann. Sequential activation of premotor, primary somatosensory and primary motor areas in humans during cued finger movements. *Clinical Neurophysiology*, 126(11):2150–2161, 2015.

[51] Albert W Marshall, Ingram Olkin, and Barry C Arnold. *Inequalities: theory of majorization and its applications*, volume 143. Springer, 1979.

[52] Gavin C Cawley and Nicola LC Talbot. On over-fitting in model selection and subsequent selection bias in performance evaluation. *Journal of Machine Learning Research*, 11(Jul):2079–2107, 2010.

[53] Camilo Toro, Günther Deuschl, Robert Thatcher, Susumu Sato, Conrad Kufta, and Mark Hallett. Event-related desynchronization and movement-related cortical potentials on the ecog and eeg. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, 93(5):380–389, 1994.

[54] Gert Pfurtscheller and Christa Neuper. Motor imagery activates primary sensorimotor area in humans. *Neuroscience letters*, 239(2-3):65–68, 1997.

[55] Steven C Cramer, Seth P Finklestein, Judith D Schaechter, George Bush, and Bruce R Rosen. Activation of distinct motor cortex regions during ipsilateral and contralateral finger movements. *Journal of neurophysiology*, 81(1):383–387, 1999.

[56] Etienne Combrisson and Karim Jerbi. Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. *Journal of neuroscience methods*, 250:126–136, 2015.

[57] Kenny Chen, Paolo Gabriel, Abdulwahab Alasfour, Chenghao Gong, Werner K Doyle, Orrin Devinsky, Daniel Friedman, Patricia Dugan, Lucia Melloni, Thomas Thesen, et al. Patient-specific pose estimation in clinical environments. *IEEE Journal of Translational Engineering in Health and Medicine*, 6:1–11, 2018.

[58] Alexander Mathis, Pranav Mamidanna, Kevin M. Cury, Taiga Abe, Venkatesh N. Murthy, Mackenzie W. Mathis, and Matthias Bethge. Deeplabcut: markerless pose estimation of user-defined body parts with deep learning. *Nature Neuroscience*, 2018.

[59] Tanmay Nath*, Alexander Mathis*, An Chi Chen, Amir Patel, Matthias Bethge, and Mackenzie W Mathis. Using deeplabcut for 3d markerless pose estimation across species and behaviors. *Nature Protocols*, 2019.

[60] Kai Joshua Miller. A library of human electrocorticographic data and analyses., 2016. data retrieved from Stanford Digital Repository, `https://purl.stanford.edu/zk881ps0522`.

[61] Manuel R Mercier, Stephan Bickel, Pierre Megevand, David M Groppe, Charles E Schroeder, Ashesh D Mehta, and Fred A Lado. Evaluation of cortical local field potential diffusion in stereotactic electro-encephalography recordings: a glimpse on white matter signal. *Neuroimage*, 147:219–232, 2017.

[62] Ali Farshchian, Juan A Gallego, Joseph P Cohen, Yoshua Bengio, Lee E Miller, and Sara A Solla. Adversarial domain adaptation for stable brain-machine interfaces. *arXiv preprint arXiv:1810.00045*, 2018.

Chapter 5

Patient-Specific Pose Estimation in Clinical Environments

Abstract

Reliable posture labels in hospital environments can augment research studies on neural correlates to natural behaviors and clinical applications that monitor patient activity. However, many existing pose estimation frameworks are not calibrated for these unpredictable settings. In this paper we propose a semi-automated approach for improving upper-body pose estimation in noisy clinical environments, whereby we adapt and build around an existing joint tracking framework to improve its robustness to environmental uncertainties. The proposed framework uses subject-specific convolutional neural network (CNN) models trained on a subset of a patient's RGB video recording chosen to maximize the feature variance of each joint. Furthermore, by compensating for scene lighting changes and by refining the predicted joint trajectories through a Kalman filter with fitted noise parameters, the extended system yields more consistent and accurate posture annotations when compared to two state-of-the-art generalized pose tracking algorithms for three hospital patients recorded in two research clinics.

5.1 Introduction

Accurate patient joint tracking and posture estimates provide quantitative data that can be experimentally and clinically informative. Upper-body annotations for long-term continuous video of patients in the epilepsy monitoring unit (EMU), for example, can be used to further explore the relationship between neural activity and unconstrained human movement when combined with a neural recording system [1, 2]. Analysis of neural correlates to behavioral labels extracted from long duration naturalistic datasets collected in the hospital could then provide a pathway for more robust brain-computer interfaces (BCI's). These include assistive robotic arms [3–5] and neural prostheses [6, 7] for those with limb loss or total paralysis. Alternatively, posture annotations can be used to objectively score patient motor capabilities to enhance current subjective assessments. For instance, the Unified Parkinson's disease rating scale (UPDRS) [8] is the current standard for evaluating the severity of motor impairment associated with Parkinson's disease, but it involves a qualitative evaluation by interview and clinical observation. The outcome of this process is limited to the clinician's interpretation during examination and can be inconsistent between evaluators. Combining such assessments with additional insight from objective motion analysis could help improve the efficacy of treatment protocols. Other motor scoring assessments (e.g., BOT-2 [9], FMA [10], MAS [11]) would benefit similarly.

Several studies in automated motor scoring incorporate wearable devices (such as inertial measurement units [12, 13], accelerometers [14–16], and internet-connected personalized healthcare systems (PHS) [17]) that collect kinematic data of subject appendages, but may risk complications from prolonged wear of physical sensors [18]. These systems can be complemented with less invasive video-based tracking methods that supplant physical sensors when they are temporarily removed for relief. Additionally, for patients who are unable to wear such sensors due to injuries at the wrists or at other attachment areas, video-based joint tracking can create a nonintrusive means to monitor their safety and well-beings.

To this end, we introduce PatientPose, an adaptation of Caffe-Heatmap [19] for semi-automated pose estimation in clinical environments. Our additions to the existing pose estimation framework include three key elements that enable more accurate and consistent patient posture tracking than before: 1) a preprocessing step to accommodate for the frequent scene lighting changes found in hospital rooms; 2) a training technique that targets separate convolutional neural network (CNN) models specifically to each patient to capture the high variance of postures a subject can realize during their hospital stay; and 3) a Kalman filter with tuned noise parameters which refines the predicted joint trajectories. We show that for three subjects recorded in two research clinics, the extended system provides an increase in tracking

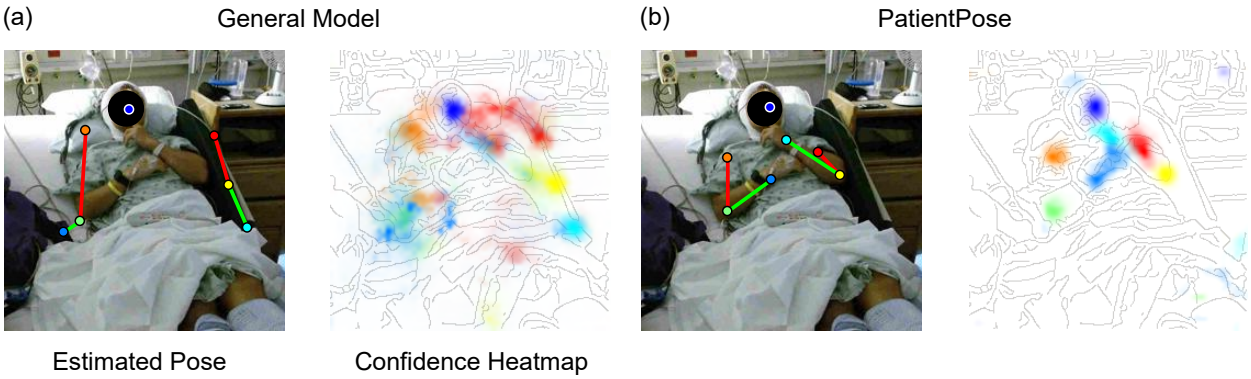


Figure 5.1 **Comparison of pose estimation models.** Upper-body posture annotations and their corresponding probability heatmaps using (a) a prepackaged model and (b) our patient-specific model. Both models were developed using the Caffe-Heatmap architecture. Our proposed framework accounts for variability in clinical environments to improve pose estimates and can be more confident and accurate than generalized methods. Subject 1 is depicted in these images.

performance when compared to two state-of-the-art generalized frameworks (Fig. 5.1).

Related Work

The importance and potential impact of human pose estimation is supported by the substantial history of research in this field. Recent work in computer vision [19–27] suggests using deep CNN's to automatically estimate joint locations in long-term recording sessions. Toshev and Szegedy [25] were the first to use CNN's for human pose estimation and regressed joint coordinates directly from a cascade of deep CNN regressors. More recently, Pfister *et al.* [19] instead regressed confidence *heatmaps* for the joint positions of each input frame and improved estimates by aligning and pooling heatmaps with neighboring frames. This framework was then extended by Charles *et al.* [26] who recursively processed the estimates for further improvements. Cao *et al.* [27] used a two-branch multi-stage CNN architecture to encode the location and orientation of body parts into a set of 2D vector fields and achieved real-time multi-person pose estimation.

While general pose estimation frameworks are effective when subjects are located in uncluttered settings, they can be unreliable when applied to noisy environments such as epilepsy monitoring and intensive care units. Such locations present several visual challenges that these generic frameworks do not account for, including variance in lighting conditions throughout a recording session, non-subject (e.g., clinician, nurse, visitor) interferences, and environmental occlusions (e.g., bed blankets, head wrapping, hospital gown). As a result, joint confidence heatmaps generated from hospital video using all-inclusive pose estimators may either be weak and distributed across the whole image, or confidently confused with

another object in the room (Fig. 5.1a).

Previous works on improving pose estimation performance in complex clinical environments take advantage of a wide range of available sensors [28–33]. Achilles *et al.* [28] used a single depth camera to regress joint coordinates specifically for body tracking under blanket occlusion, and Liu *et al.* [29] relied on a novel infrared image acquisition technique using a bird's-eye view in order to monitor patient sleeping postures. Belagiannis *et al.* [30] combined information from multiple RGB cameras to track surgeons and medical staff in operating rooms, and Kadkhodamohammadi *et al.* [31] improved upon pose estimation in operating rooms by using depth sensors in tandem with multiple RGB cameras. Chaaraoui *et al.* [32] also used a multi-camera setup but for vision-based monitoring and action recognition by learning subject activity patterns from estimated silhouettes. However, none have attempted to extract high-quality joint estimates to track freely-behaving patients in hospitals across hours of data using a single RGB camera. Capturing RGB video is trivial with the current state of consumer technology, and to our knowledge this work is the first to create a pose estimation framework that specifically targets subjects in clinical environments using only one angle of recorded RGB video. Additionally, the proposed extensions to Pfister *et al.*'s Caffe-Heatmap [19] do not modify the original framework's central CNN architecture and could potentially be adopted to improve other general pose estimators (Fig. 5.2), and our framework is capable of a real-time implementation after a patient's initial training procedure.

5.2 Methods and Procedures

5.2.1 Subject Recording and Dataset Description

In this study, we conducted our experiments using a novel dataset. Three patients with intractable epilepsy were enrolled according to protocols approved by the Institutional Review Board (IRB) at the New York University (NYU) Langone Comprehensive Epilepsy Center and the Rady Children's Hospital (RCH), San Diego, Pediatric Epilepsy Center. Video was recorded using a Microsoft Kinect v2 during each patient's stay, targeting 1–2 days post-implant of electrodes when the subjects were expected to be most active. Video was recorded using multiple modalities (i.e., RGB, depth, infrared), but only the RGB images were considered for this study. Specific details regarding the duration of each subject's recording session and the number of frames used for framework training/evaluation are provided in Table 5.1. Note that the Kinect v2 RGB camera samples at either 15 or 30 frames-per-second (fps) depending on room luminance and horizontally flips all images when saving to disk. Our data acquisition system was fit onto a custom-built mount that stood five feet tall and was placed about 20 degrees to the left of Subjects 1 and 3 (S1 and

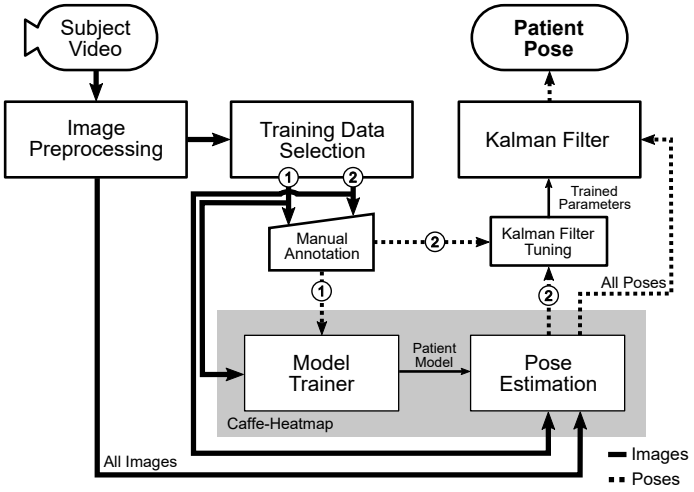


Figure 5.2 **Pipeline of proposed framework.** The proposed framework extends Caffe-Heatmap to improve pose estimation of patient video recorded in clinical environments. Prior to estimation, a new patient-specific CNN model is trained using a subset of preprocessed video frames that maximizes feature variance (①). This model is then used to estimate the joint positions of the same patient from additional video, which are then refined using a Kalman filter with noise parameters trained using another subset of preprocessed frames (②). This work used 2,000 frames for ① and 500 frames for ②.

S3) and 45 degrees to the left of Subject 2 (S2).

5.2.2 Image Preprocessing

Cropping

To maintain memory efficiency during GPU training, recorded RGB frames were cropped and re-sized from 1920x1080 to 256x256 pixels in width and height. The location of cropping was centered around the patient and manually selected once per patient dataset.

Scene Lighting Normalization

To account for the fluctuations in lighting conditions often found in hospital rooms, image brightness was normalized by first transforming each frame to the Hue-Saturation-Value (HSV) color space and then applying contrast-limited adaptive histogram equalization (CLAHE) [34] onto the value layer with an 8x8 tile size. Regions with similar surroundings (e.g., bed sheets) were susceptible to noise amplification when normalized using global or regular adaptive equalization [35], and CLAHE limited the amount those regions could increase in contrast (Fig. 5.3).

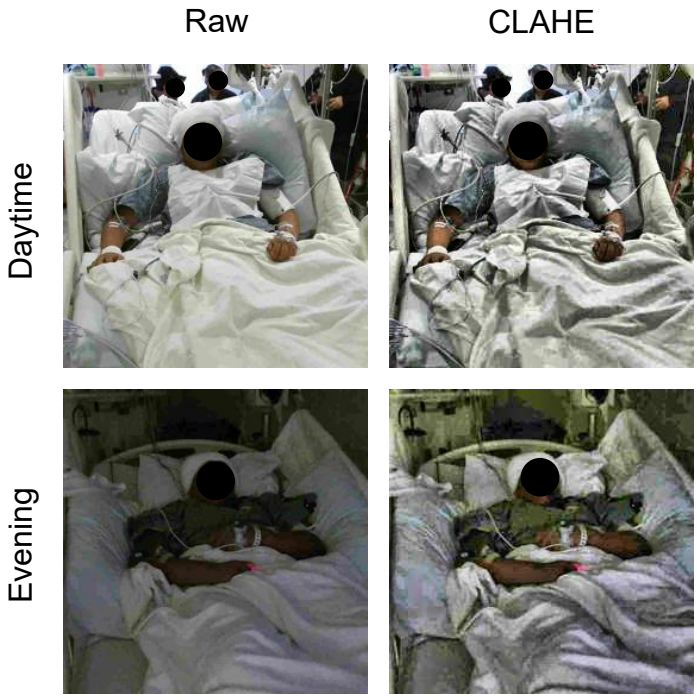


Figure 5.3 **Scene lighting normalization.** Raw patient images (left) during daytime (top) and evening (bottom) were significantly different in lighting conditions for the same clinic. However, after applying contrast-limited histogram equalization (CLAHE, right), the frames across a dataset became more consistent in brightness. Subject 3 is depicted in these images.

5.2.3 Convolutional Neural Network Models

Motivation

Convolutional neural networks can be used to build models that predict subsequent data by learning and extracting patterns from a training set; this machine learning technique is used in a wide range of applications aside from pose estimation, such as human action recognition [36], predicting blood glucose levels [37], natural language processing [38], and more [39–42]. However, the performance of a trained model is heavily reliant on the quality of its training data. In human pose estimation, prepackaged CNN models trained using movie or video frames work well against other generic pose estimation datasets (e.g., British Broadcasting Corporation (BBC) Pose [43], Common Objects in Context (COCO) [44], Frames Labeled in Cinema (FLIC) [45], Max Planck Institute for Informatics (MPII) Human Pose [46]), but can be less reliable when applied to videos of hospital patients due to various challenges unique to the clinical setting. Therefore, we trained a separate CNN model for each of our subjects using an extracted subset of frames held out from the test set (Fig. 5.2). These high-quality training sets were designed to capture the wide range of postures the corresponding patient may naturally take on throughout a recording session.

Extracting High-Quality Training Data

To maximize posture diversity and therefore feature variance in a patient's training data, frames were selected from both movement and non-movement periods. This was accomplished by first applying the Gunnar-Farnebäck dense optical flow algorithm [47] onto the raw RGB video of the same patient to calculate the average magnitude of scene movement between adjacent frames. A threshold on this average flow empirically set to 0.15 pixels per frame then partitioned patient RGB video into periods of movement and idleness. Afterwards, a subset of frames was uniformly sampled from the segmented video such that frames drawn from movement and rest periods were distributed $70\%/30\%$. Using this strategy, 2,000 frames for model training were selected across the entire span of each patient's dataset which captured different postures the patient may take on during their stay. Frames with significant patient occlusions were manually excluded.

Patient-Specific Model Training

Ground truth (x,y) coordinates of the seven joints (i.e., head, left/right hands, elbows, and shoulders) were manually marked for each training set using a custom labeling script. A CNN model was then trained for each patient using the Caffe-Heatmap model training architecture [19] on an NVIDIA GeForce GTX 1080 Ti GPU with the annotated images. One million iterations of batch size 14 were used with a learning rate of 10^{-8} and momentum of 0.95, and each iteration took approximately 0.75 seconds for a total of nine days of training per model; these hyperparameter values were chosen to match those of the original Caffe-Heatmap. The resulting models learned features specific to each patient through the high-quality training set. Using the same hardware and configurations, training a generic model on the FLIC dataset with about 4,500 frames would span around twelve days.

5.2.4 Inference via Patient-Specific Model

To enable easy adoption of our augmentations onto other pose estimators, we did not directly modify the Caffe-Heatmap framework. We therefore treated it as a black box, with the inputs as the patient-specific Caffe [48] model and N number of frames, and an output of seven 256x256 confidence heatmaps for each frame. Each joint location was then taken to be at the *argmax* of its corresponding heatmap, resulting in a 2x7xN structure of $(x,y) \in [0,256]$ joint coordinates. For each frame, inference spanned $\sim$0.03 seconds when using the same NVIDIA GeForce GTX 1080 Ti (compared to $\sim$10 seconds per frame on an Intel Xeon CPU E5-2630), enabling the potential for a real-time implementation. Specific details of the

Caffe-Heatmap architecture can be found in [19].

5.2.5 Kalman Filter

Motivation

Joint locations estimated by a patient-specific CNN model were generally reasonable, but occasionally contained jitter or large jumps when a patient moved quickly or was occluded. Therefore, a standard Kalman filter [49] was used as a post-processing step to leverage the temporal information found between frames in order to refine any noisy measurements. The Kalman filter consists of two primary components (a state transition function and a measurement function) that model the underlying physics of a system to predict its state over time, making it an appropriate choice for denoising estimated joint trajectories. In addition, we chose to use a Kalman filter (as opposed to a non-causal Kalman smoother [50]) to preserve our framework's potential to be implemented in real-time due to the filter's causality.

General Equations

The Kalman filter is a recursive two-step process which iteratively predicts a system's next state using past information and a predefined model, then updates its predictions using external sensor measurements. These two functions are defined by the linear state transition and measurement matrices A and H. In addition, the estimated state μ_t at each t^{th} iteration is accompanied with a covariance Σ_t that measures the accuracy of the estimate at that time step. In the Kalman filter's prediction step, we have:

$$\hat{\mu}_t = A\mu_{t-1}\,, \tag{5.1}$$

$$\hat{\Sigma}_t = A\Sigma_{t-1}A^\top + Q\,, \tag{5.2}$$

where the hat indicates that these values are purely estimates by the filter without considering any outside measurements yet. The Q term above is the covariance of the process noise that captures the error between the transition model and the true dynamics of the system, and it is assumed to be Gaussian distributed in this work. In the update step, we have:

$$K_t = \hat{\Sigma}_t H^\top (H\hat{\Sigma}_t H^\top + R)^{-1}\,, \tag{5.3}$$

$$\mu_t = \hat{\mu}_t + K_t(z_t - H\hat{\mu}_t)\,, \tag{5.4}$$

$$\Sigma_t = \hat{\Sigma}_t - K_t H\hat{\Sigma}_t\,, \tag{5.5}$$

where K_t is the Kalman gain that adjusts the next predicted state μ_t and covariance Σ_t depending on the accuracy of the model. In this step, external sensor measurements z_t provide the filter with additional

information on the system's next possible state, and the R term captures the noise in these measurements (also assumed to be Gaussian). Complete derivations of these equations can be found in [51–53].

In this work, we used a constant velocity model [54] to define the state transition and measurement matricies A and H in these equations, and we assumed independent movement between the seven joints. Therefore, we ran a separate Kalman filter on each joint, in which the 4D state estimates μ_t contained a joint's (x,y) pixel position and velocity at time t. Additionally, we used the (x,y) coordinates provided by a patient's CNN model as the external z_t measurements to update the filter's predictions on the system's state. These Kalman filter equations recursively computed a next-best-guess on a joint's position using the predefined constant velocity model and the pose estimates by the CNN model.

Learning the Noise Parameters

The Q and R process and measurement noise covariances are critical components to the Kalman filter that model unforeseen perturbations on the system. In the context of this work, the Q term captures how erroneous the constant velocity model is to the real dynamics of a patient and the R term captures the variability in the CNN's pose estimates to the true positions. However, these matrices are frequently difficult to estimate and are often constructed using prior knowledge of the problem, tediously tuned by hand, or assumed to be independent between variables for convenience. Abbeel *et al.* [55] demonstrated that Q and R can be learned by maximizing the joint likelihood between the states and the measurements of a training dataset. More specifically, for T number of training datapoints, the optimal parameters Q^j_{MLE} and R^j_{MLE} for each j^{th} patient joint can be formulated as:

$$\left\langle Q^j_{MLE}, R^j_{MLE} \right\rangle = \arg\max_{Q,R} \log p(x^j_{0:T}, z^j_{0:T}) \,. \tag{5.6}$$

Here, the joint probability distribution between the sequence of ground truth states $x^j_{0:T}$ and CNN pose estimates $z^j_{0:T}$ is:

$$p(x^j_{0:T}, z^j_{0:T}) = p(x^j_0) \prod_{t=1}^{T} p(x^j_t \,|\, x^j_{t-1}) \prod_{t=0}^{T} p(z^j_t \,|\, x^j_t) \tag{5.7}$$

with some prior $p(x^j_0)$, where the Gaussian motion and observation models are:

$$p(x^j_t \,|\, x^j_{t-1}) = \mathcal{N}(x^j_t; A x^j_{t-1}, Q) \,, \tag{5.8}$$

$$p(z^j_t \,|\, x^j_t) = \mathcal{N}(z^j_t; H x^j_t, R) \,. \tag{5.9}$$

Substituting (5.7), (5.8), and (5.9) into (5.6) and then computing the closed form solutions results in the

equations:

$$Q_{MLE}^{j} = \frac{1}{T}\sum_{t=1}^{T}(x_t^j - Ax_{t-1}^j)(x_t^j - Ax_{t-1}^j)^\top, \tag{5.10}$$

$$R_{MLE}^{j} = \frac{1}{T+1}\sum_{t=0}^{T}(z_t^j - Hx_t^j)(z_t^j - Hx_t^j)^\top. \tag{5.11}$$

Patient-Specific Noise Parameter Training

Equations (5.10) and (5.11) require ground truth states x_t^j and CNN pose estimates z_t^j from a set of training data for each patient joint. In addition, because (5.10) depends on states at times t and $t-1$, the training joints must be continuous over time. Therefore, in an attempt to capture the process variability across the entire span of a patient dataset, we constructed a "semi-continuous" training subset using the following steps. First, we segmented a patient's video into periods of movement and idleness using the same optical flow method as described in *Section II.C.1*. Afterwards, we extracted the first 10 frames of a movement period for 50 periods chosen uniformly across the span of the patient's video. This resulted in a set of 500 "semi-continuous" frames (50 discontinuous movement periods of 10 continuous frames each) for each patient which we used to train patient-specific noise parameters. Occluded segments, movements less than 10 frames in length, and frames used for evaluation were excluded during period selection. After extraction, these 500 frames were manually annotated for ground truth joint positions. To obtain the ground truth states x_t^j, joints were assumed to have zero initial velocity at the start of each movement period, and the remaining velocity values were calculated as the difference in pixel position between adjacent frames within the same period. The frames of each segment were then sent through Caffe-Heatmap's pose estimator along with the corresponding patient-specific model to obtain z_t^j.

To calculate a patient's set of measurement noise covariances R_*, we directly implemented (5.11) for each joint such that $R_*^j = R_{MLE}^j$ using all 500 training datapoints. Equation (5.11) only depends on values at time t, and therefore its training set need not be continuous. However, because (5.10) depends on values at times t and $t-1$, we first calculated a separate $Q_{MLE}^{j,m}$ for each m^{th} movement period of length $T = 10$ frames in the semi-continuous set using (5.10). The resulting $M = 50$ matrices per joint were the covariances that maximized the data likelihood in their corresponding movement sequence. A joint's process noise parameter Q_*^j was then taken to be the average of these covariances, such that:

$$Q_*^j = \frac{1}{M}\sum_{m=1}^{M} Q_{MLE}^{j,m}. \tag{5.12}$$

These calculated parameters $Q_* \in \mathbb{R}^{4x4x7}$ and $R_* \in \mathbb{R}^{2x2x7}$ for each patient modeled any unforeseen perturbations on the system throughout a patient's dataset at runtime of the filter.

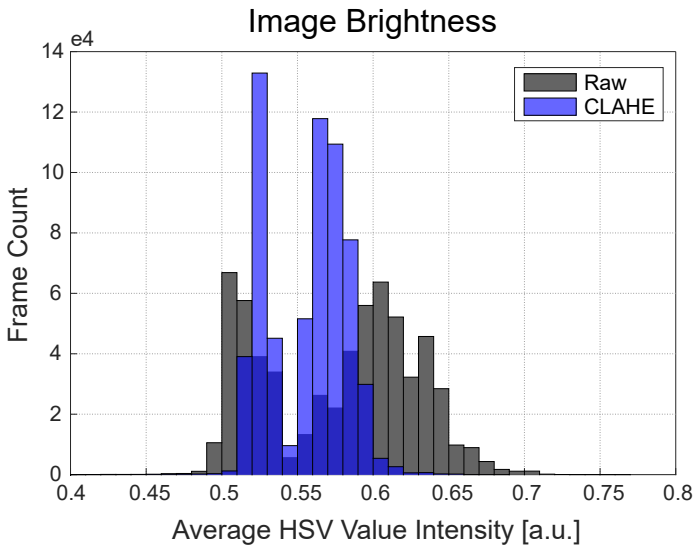


Figure 5.4 **Mean frame brightness before and after CLAHE normalization.** Overlaid distributions of the average image brightness before (gray) and after (blue) CLAHE confirm that the lighting conditions across images are more similar after equalization. Average brightness of a frame was measured by taking the mean of a frame's value-layer intensity. Each histogram used the entire Subject 2 dataset (N = 625,127 frames).

Table 5.1 Dataset Summary

Subject	**Study ID**	**Hours**	**Number of Frames**		
			Total	*Training*	*Testing*
S1	NY531	1.8	94,470	2,500	3,000
S2	RCH1	5.8	625,127	2,500	1,000
S3	RCH3	22.2	2,399,469	2,500	1,000

5.3 Results

In this section, we first provide an analysis of each component to convince the reader that our additions to the original framework are reasonable for improving pose estimation in clinical environments. Then, to validate our methods as a whole, we present pose estimation accuracy comparisons between our framework and two state-of-the-art generalized frameworks using selected test sets of patient data for three subjects recorded in various clinical monitoring units. A representative demo video of patient pose estimation can be viewed at https://youtu.be/c3DZ5ojPa9k.[1]

5.3.1 Analysis of Components

Scene Lighting Normalization

After normalizing image brightness by applying CLAHE onto the value layer of each HSV frame, we observed a significant reduction in scene lighting variance throughout each patient dataset. This reduction is depicted by the histograms of the mean V-channel magnitude for each frame in the S2 dataset using 0.005 bin widths before and after lighting normalization (Fig. 5.4). The value-layer in the HSV color space corresponds to image brightness, and therefore the lower histogram variance after normalization indicates a higher similarity in lighting conditions within the patient dataset than before. This translates into joint features that are more likely to be consistent in visibility.

High-Quality Training

To establish that our high-quality training strategy can capture a large variety of postures within a patient dataset, we compared against another manually annotated subset defined as the first 15 minutes of frames for the same patient ($\sim$13.5k frames at 15 fps). Patients were observed to engage in different postures depending on the time of day (e.g., upright vs. rest), and we therefore inferred that frames extracted using our "high-variance" (HV) training strategy would contain greater posture diversity than frames within this "low-variance" (LV) set. To investigate this, we first defined each posture as a 14-dimensional vector containing the (x, y) pixel coordinates for each of the seven joints. These vectors were then projected down to 2-dimensional space using t-distributed stochastic neighbor embedding (t-SNE) [56] for a graphical intuition of the posture coverage between the two strategies. Only unique datapoints were considered in this analysis, and we observed that the HV set initially contained twice as many unique postures than the LV set. Therefore, prior to t-SNE dimensionality reduction, we uniformly sampled the HV set to ensure an equal

[1] Video frames have been blurred for patient confidentiality

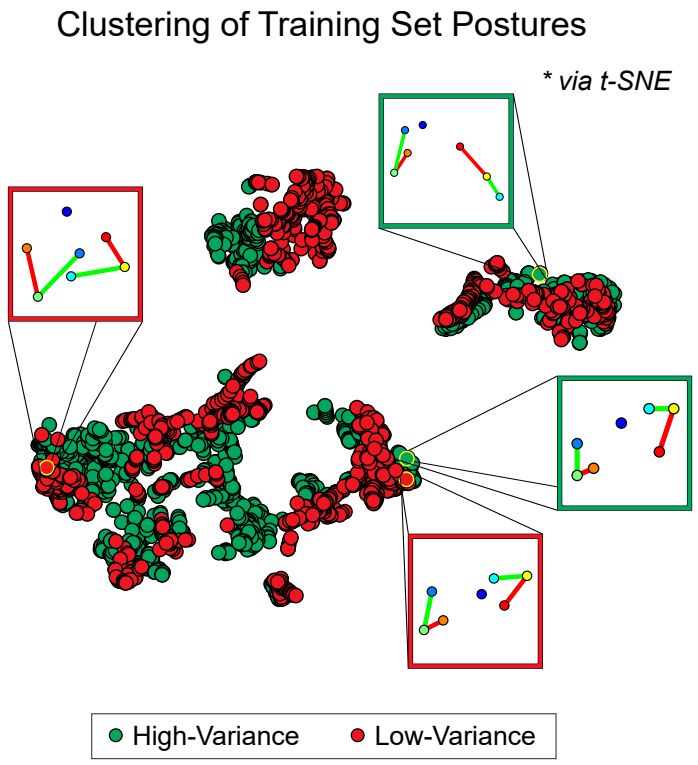


Figure 5.5 **Visualization of posture varieties covered by training strategies.** Manually annotated postures from two training strategies were projected onto a 2D space using t-SNE to provide a graphical intuition of the various poses included in each set. ”High-variance” training frames were selected from periods of movement and rest across the entire span of the patient dataset, whereas ”low-variance” training frames were the first 15 minutes of recording. Note that the ”high-variance” set initially contained twice as many unique postures than ”low-variance” but was uniformly downsampled to prevent bias in the projection. Points within the same cluster resemble similar postures.

number of datapoints that would have otherwise biased the t-SNE manifold towards the more represented HV postures. In addition, the two sets were concatenated prior to projection to ensure compatibility in the output space. In this analysis, the exact t-SNE algorithm was implemented with a standard Euclidean distance metric for 1,000 iterations at a perplexity of 50 and a learning rate of 500, and the two subsets were derived from S1.

The results after projecting the 14D postures onto a 2D space (Fig. 5.5) represent a low-dimensional clustering of different patient postures extracted from the two training strategies. Each color-coded datapoint represents a unique set of seven joint coordinates, and points within the same cluster resemble similar postures. Despite downsampling the HV set to match the size of the LV set for an unbiased projection, the HV set still visually occupies a larger area in the projected space. This suggests that our high-quality training strategy can capture a diverse collection of patient postures. In addition, the spread of the HV datapoints encompasses nearly all of the LV points, indicating that there may be little to no trade-off between posture diversity and coverage quality when extracting training frames from the entire span of video. Therefore, our high-quality training strategy constructs a more informative training set when compared to frames extracted from a limited window of time and can provide the CNN architecture with more representations of each joint to train on.

Kalman Filter with Trained Noise Parameters

Immediate joint coordinates estimated by Caffe-Heatmap using a patient-specific CNN model are still subject to inconsistencies during periods of quick movements or patient occlusions. However, a Kalman filter with trained noise parameters refines these predictions and reduces the jitter and noise within estimated paths. Prior to optimizing S1's left hand in the testing data, the original trajectory demonstrated reasonable tracking with an average error of 10.42 ± 5.85 pixels from the ground truth. In contrast, a denoised trajectory using patient-specific parameters followed the true path more closely at an average error of 8.23 ± 5.19 pixels and exhibited less jitter at sharp turns. This is illustrated in Fig. 5.6 which shows a segment of S1's left hand trajectory using the different Q and R noise parameters. For reference, using stock constant velocity parameters resulted in an average error of 9.94 ± 6.34 pixels within the same test set. These observations were consistent throughout all joints and subjects.

5.3.2 Pose Estimation Results

Performance was measured using the Euclidean distance of estimated joint coordinates against an additional set of manually annotated frames held out from the training set for each patient. These frames

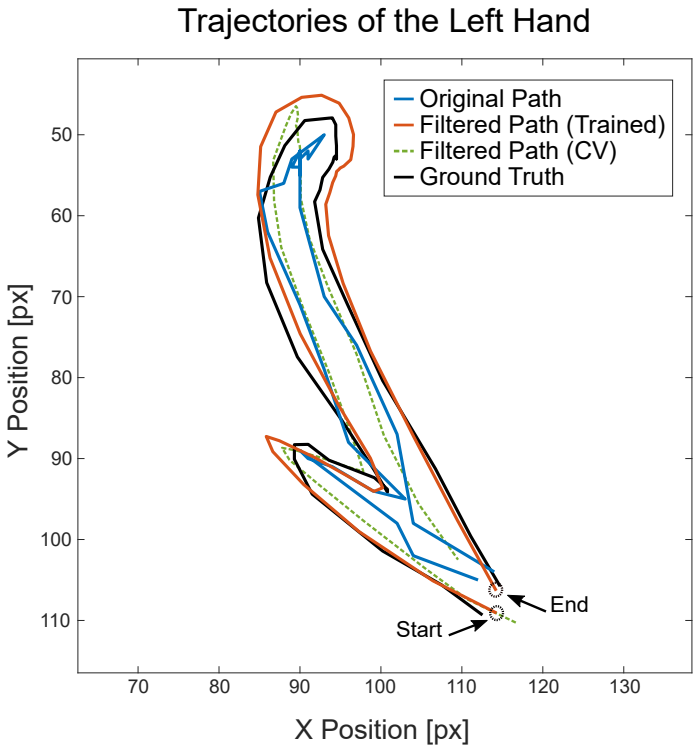


Figure 5.6 **Comparison of trajectories.** Estimated (x, y) coordinates of S1's left hand during an example segment of movement before (blue) and after (orange) using a Kalman filter with fitted noise parameters, as compared to the ground truth (black). A filtered path using constant velocity (CV) noise parameters (green) is also provided for reference. Across all testing data for Subject 1's left hand, the trained Kalman filter produced a lower average prediction error of 8.23 $\pm$5.19 pixels from the ground truth, compared to the original path's error of 10.42 $\pm$5.85 pixels.

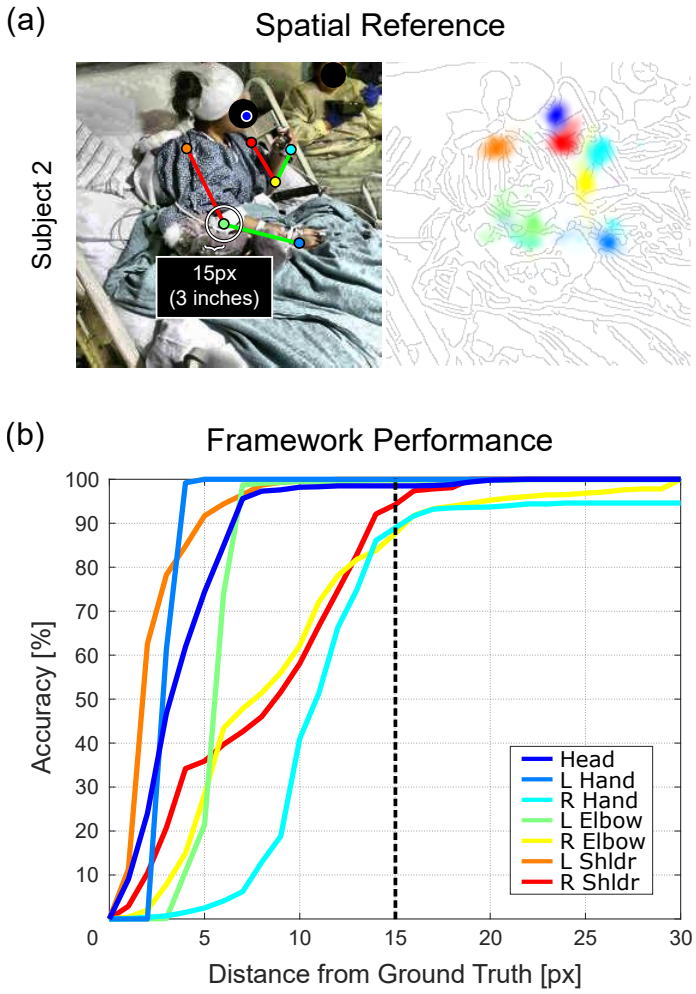


Figure 5.7 **Spatial reference and S2 performance.** (a) Example skeleton and heatmap with a 15-pixel (3-inch) radius for spatial reference, and (b) Subject 2 accuracy curves between 0 and 30 pixel tolerances from the ground truth.

were chosen for their variety in postures, fluctuations in lighting conditions, and occasional nurse appearances. For each patient test set, we compared our framework's pose estimation performance against two state-of-the-art generic frameworks by evaluating joint estimates from each method at distances between 0 and 30 pixels (px) from the ground truth. These methods included Caffe-Heatmap by Pfister *et al.* [19] trained on FLIC and OpenPose by Cao *et al.* [27] trained on COCO.[2] Fig. 5.7a provides a spatial reference of 15 pixels (approximately 3 inches) and Fig. 5.7b shows the joint accuracies at varying tolerances for Subject 2's test set. For a progression of pose estimation performance after each proposed contribution, refer to Fig. 5.9 in the appendix.

At a tolerance of 15 pixels, our framework was more accurate than Caffe-Heatmap by $42.4 \pm 8.3\%$

[2]Models were provided out-of-box by their respective authors

and OpenPose by $11.4 \pm 3.9\%$ on average across our three patient test sets (Table 5.2). Patient hands and elbows were typically the most challenging joints to estimate for every method, but we saw more consistent tracking in these categories using our framework. Fig. 5.8 shows a complete performance comparison against the two generalized methods for all three subjects. With Subjects 1 and 2, we observed a considerable improvement on tracking performance for all seven joints, and our framework labeled at least 80% of frames for any joint within 15 pixels from the ground truth. In addition, our framework provided $\sim$50% more hand annotations at this tolerance when compared to OpenPose for these two subjects.

In contrast to the test sets for Subjects 1 and 2, Subject 3's chosen test set contained more frequent hand occlusions in which Subject 3 often placed their left and right hands behind their head during rest. This decreased the overall tracking consistency across all methods for these two joints. However, our framework still on average provided $22.0 \pm 9.5\%$ and $9.8 \pm 6.8\%$ more hand labels than Caffe-Heatmap and OpenPose, respectively. For Subject 3's elbows, the second most challenging category, we saw an overall increase in performance by $38.2 \pm 17.7\%$ against Caffe-Heatmap and $11.3 \pm 5.7\%$ against OpenPose when using our framework. This suggests that our framework can be more consistent within reasonable spatial tolerances for particularly noisy segments of video compared to general methods.

5.4 Conclusion

In this paper, we presented several extensions onto an existing pose estimation framework to improve posture tracking in clinical environments. By extracting images from periods of movement and idleness across the entire span of a patient's dataset, we can construct a subset of training frames that captures a diverse collection of postures for a patient-specific CNN model. Furthermore, by accounting for the frequent lighting changes often found in these environments and by refining the predicted trajectories through a Kalman filter with trained noise parameters, our framework can provide more reliable annotations on a patient's pose in these settings when compared to generic pose estimation frameworks.

Our framework relies solely on low-resolution RGB images to be implemented and therefore can be used by anyone with a means of recording RGB video. In addition, our augmentations can be potentially adopted to improve other pose estimators, and our framework is capable of running in real-time after training as a consequence of the Kalman filter's causality. We have open-sourced our standalone PatientPose toolbox,[3] and we encourage others to use our framework for their own experimental or clinical studies or to apply and build upon our methods. However, we suggest that the trade-off between PatientPose and

[3] https://github.com/TNEL-UCSD/PatientPose

Table 5.2 Pose Estimation Accuracy Rates @ 15px [%]

Subject 1

Method	Head	Hands	Elbows	Shoulders	Average
CH-FLIC	95.6	0.4	22.1	30.7	49.8 ±41.0
OpenPose	99.4	37.3	93.4	88.4	83.7 ±28.6
Ours	**99.9**	**87.8**	**99.1**	**95.6**	**96.5 ±5.6**

Subject 2

Method	Head	Hands	Elbows	Shoulders	Average
CH-FLIC	78.4	2.0	16.6	48.8	49.2 ±34.1
OpenPose	**99.4**	49.4	69.2	92.9	82.3 ±23.1
Ours	98.5	**94.5**	**93.7**	**97.1**	**96.8 ±2.2**

Subject 3

Method	Head	Hands	Elbows	Shoulders	Average
CH-FLIC	82.0	38.9	23.9	12.0	49.7 ±30.9
OpenPose	97.9	48.0	52.5	79.5	75.6 ±23.5
Ours	**99.4**	**60.1**	**73.0**	**80.2**	**82.5 ±16.4**

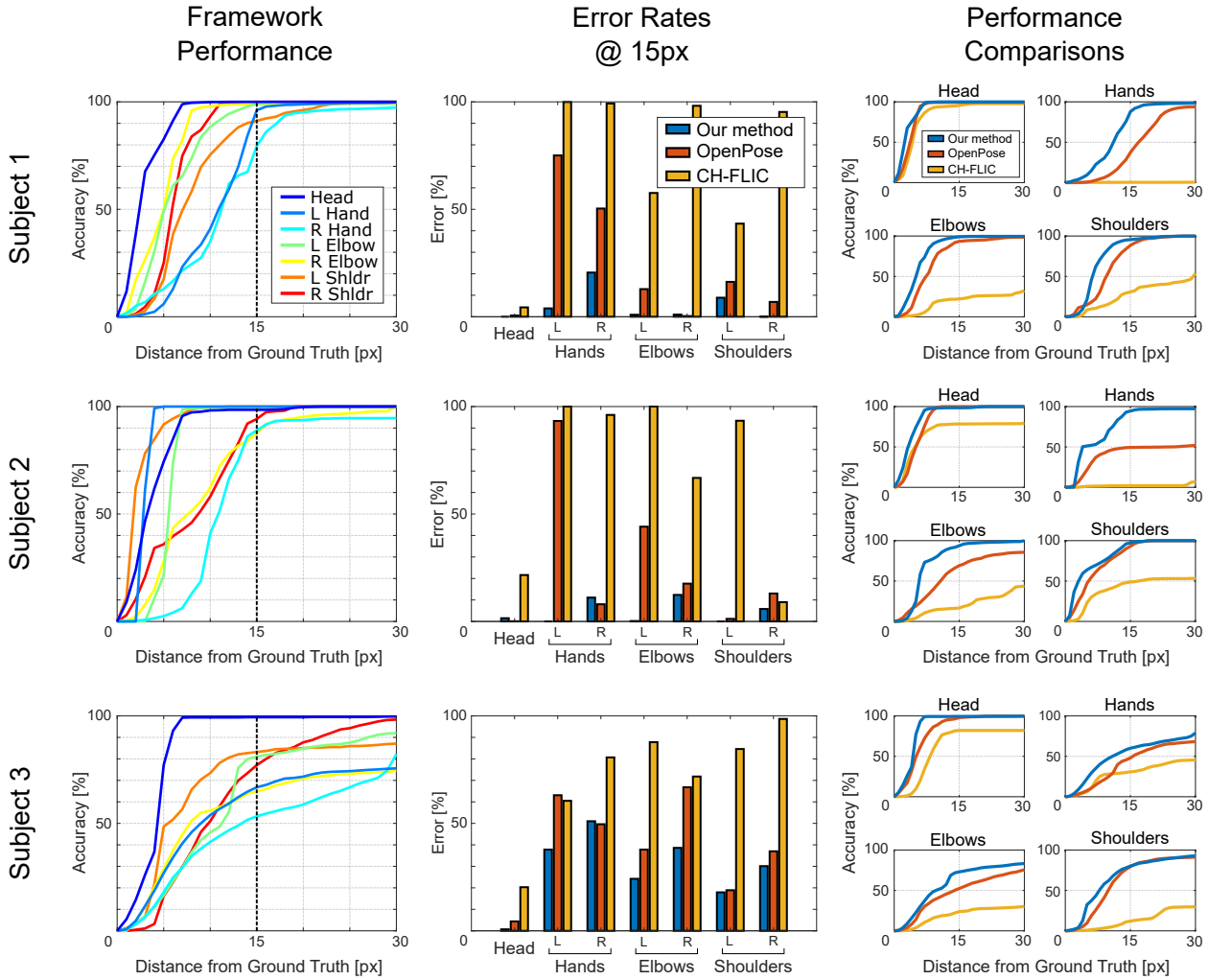


Figure 5.8 **Pose estimation comparison.** Performance of our proposed method as compared to OpenPose and Caffe-Heatmap with FLIC (CH-FLIC) for each subject is shown above. The left column provides the accuracies of each joint at various tolerances from the ground truth using our framework, and the middle compares the error rates at 15 pixels. The right column compares the accuracies of each category of joints averaged between the left and right body parts.

generalized frameworks should be considered before use. In particular, although we have demonstrated the potential to substantially improve posture estimation quality with our add-ons, we note that our framework's upfront cost of labeling and training a separate CNN model for each patient is greater. Frameworks that are built independent from subjects and environments are often prepackaged with general models that can be applied to patient data right away without any additional work, and therefore may be the preferred choice for those seeking an immediate solution. However, for others who require a more custom approach which can result in a higher consistency and accuracy of pose annotations in these environments, we encourage them to look into PatientPose as a means to extend beyond current general methods.

This trade-off directly motivates future work that could explore the use of insights from generalized frameworks in order to reduce the upfront efforts per patient, or to develop a framework for "hospital-specific" models that generalize across patients within the same hospital using techniques such as transfer learning. Specifically, a reduction in training time can be achieved by using more powerful hardware and software solutions as this area of research and development continues to mature, and a "hospital-specific" framework which meets halfway between general and specific methods could mitigate concerns of model overfitting in more varied clinical environments. Work in this field will continue to expand with the increasing desire for automated behavioral labels, since these labels can be informative for both research studies and clinical applications. Analysis of neural correlates to natural behaviors extracted from pose estimates, for example, could enable more robust brain-machine prostheses that would benefit those with motor disabilities. In addition, patient tracking can provide a way to automate patient safety monitoring and could improve current motor scoring assessments, overall patient management, and the effectiveness of treatment protocols. Such studies and applications all seek to improve the quality of our health care.

Acknowledgments

We would like to thank those at UC San Diego, Rady Children's Hospital of San Diego, and the Comprehensive Epilepsy Center at the New York University Langone Medical Center. We specifically thank Preet Minas and Hugh Wang for their contributions to data collection and annotation.

This chapter is a reprint of the material as it appears in Chen, Kenny, Paolo Gabriel, Abdulwahab Alasfour, Chenghao Gong, Werner K. Doyle, Orrin Devinsky, Daniel Friedman, Patricia Dugan, Lucia Melloni, Thomas Thesen, David Gonda, Shifteh Sattar, Sonya Wang, and Vikash Gilja. "Patient-specific pose estimation in clinical environments." *IEEE Journal of Translational Engineering in Health and Medicine* 6 (2018): 1-11.

5.5 Supplement

A progression of pose estimation performance after each contribution to the original framework is shown in Fig. 5.9. These observations were consistent across all subject test sets.

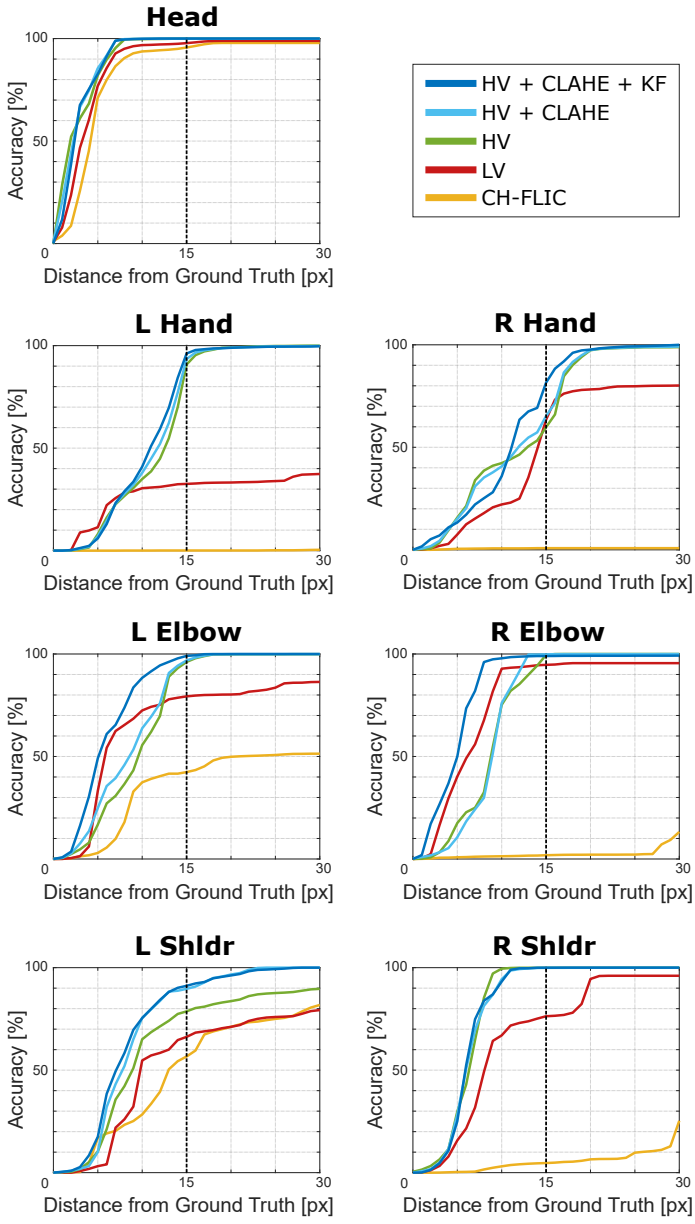


Figure 5.9 **Step-wise performance.** A progression of pose estimation performance after each contribution for all joints is shown above, comparing combinations of high-variance training (HV), lighting normalization (CLAHE), and Kalman filtering (KF) to low-variance training (LV) and Caffe-Heatmap with FLIC (CH-FLIC). The results for Subject 1's test set are shown here.

References

[1] Paolo Gabriel, Werner K. Doyle, Orrin Devinsky, Daniel Friedman, Thomas Thesen, and Vikash Gilja. Neural correlates to automatic behavior estimations from RGB-D video in epilepsy unit. In *Proceedings of the International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2016.

[2] Nancy Xin Ru Wang, Ali Farhadi, Rajesh Rao, and Bingni Brunton. AJILE movement prediction: Multimodal deep learning for natural human neural recordings and video. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence*, 2018.

[3] Carmena J, Lebedev M, Crist R, O'Doherty J, Santucci D, Dimitrov D, Patil P, Henriquez C, and Nicolelis M. Learning to control a brain–machine interface for reaching and grasping by primates. *PLoS Biology*, **1**(2), 10 2003.

[4] Velliste M, Perel S, Spalding C, Whitford A, and Schwartz A. Cortical control of a prosthetic arm for self-feeding. *Nature*, **458**:1098–1101, 2008.

[5] John K. Chapin, Karen A. Moxon, Ronald S. Markowitz, and Miguel A. L. Nicolelis. Real-time control of a robot arm using simultaneously recorded neurons in the motor cortex. *Nature Neuroscience*, 2:664–670, 1999.

[6] Gilja V, Nuyujukian P, Chestek C, Cunningham J, Krishna S, et al. A high-performance neural prosthesis enabled by control algorithm design. *Nature Neuroscience*, **15**:1752–1758, 2012.

[7] Hochberg L, Serruya M, Friehs G, Mukand J, Saleh M, et al. Neuronal ensemble control of prosthetic devices by a human with tetraplegia. *Nature*, **442**:164–171, 2006.

[8] Christopher C. Goetz. The Unified Parkinson's Disease Rating Scale (UPDRS): Status and recommendations. *Movement Disorders*, 18, 2003.

[9] Jean Crosetto Deitz, Deborah Kartin, and Kay Kopp. Review of the Bruininks-Oseretsky test of motor proficiency, second edition (BOT-2). *Physical & Occupational Therapy In Pediatrics*, 27:87–102, 2007.

[10] Axel R Fugl-Meyer, Lisbeth Jaasko, Ingegerd Leyman, Sigyn Olsson, and Solveig Steglind. The post-stroke hemiplegic patient: A method for evaluation of physical performance. *Scandinavian Journal of Rehabilitation Medicine*, 7:13–31, 1975.

[11] Janet H. Carr, Roberta B. Shepherd, Lena Nordholm, and Denise Lynne. Investigation of a new motor assessment scale for stroke patients. *Physical Therapy*, 65(2):175–180, 1985.

[12] Christina Strohrmann, Rob Labruyere, Corinna N. Gerber, Hubertus J. van Hedel, Bert Arnrich, and Gerhard Troster. Monitoring motor capacity changes of children during rehabilitation using body-worn sensors. *Journal of Neuroengineering and Rehabilitation*, 10:83, jul 2013.

[13] Avinash Parnandi, Eric Wade, and Maja J. Matarić. Motor function assessment using wearable inertial sensors. *Proceedings of the International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, aug 2010.

[14] Dheeraj Kumar, Jayavardhana Gubbi, Bernard Yan, and Marimuthu Palaniswami. Motor recovery monitoring in post acute stroke patients using wireless accelerometer and cross-correlation. *Proceedings of the International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2013.

[15] Jayavardhana Gubbi, Aravinda S. Rao, Kun Fang, Bernard Yan, and Marimuthu Palaniswami. Motor recovery monitoring using acceleration measurements in post acute stroke patients. *BioMedical Engineering OnLine*, 12, apr 2013.

[16] Alfredo Lucas, John Hermiz, Jamie LaBuzetta, Yevgeniy Arabadzhi, Navaz Karanjia, and Vikash Gilja. Use of accelerometry to monitor motor impairment to unilaterally impaired stroke patients. *Unpublished*.

[17] Jun Qi, Po Yang, Geyong Min, Oliver Amft, Feng Dong, and Lida Xu. Advanced internet of things for personalised healthcare systems: A survey. *Pervasive and Mobile Computing*, 41:132–149, 2017.

[18] Michael Schukat, David McCaldin, Kejia Wang, Guenter Schreier, Nigel H. Lovell, Michael Marschollek, and Stephen J. Redmond. Unintended consequences of wearable sensor use in healthcare. *International Medical Informatics Association (IMIA) Yearbook of Medical Informatics*, pages 73–86, 2016.

[19] Tomas Pfister, James Charles, and Andrew Zisserman. Flowing convnets for human pose estimation in videos. In *Proceedings of the IEEE International Conference on Computer Vision*, 2015.

[20] Rob Fergus, George Williams, Ian Spiro, Christoph Bregler, and Graham W. Taylor. Pose-sensitive embedding by nonlinear NCA regression. In *Proceedings of the Advances in Neural Information Processing Systems (NIPS)*, 2010.

[21] Arjun Jain, Jonathan Tompson, Mykhaylo Andriluka, Graham W. Taylor, and Christoph Bregler. Learning human pose estimation features with convolutional networks. In *Proceedings of the International Conference on Learning Representations*, 2014.

[22] Arjun Jain, Jonathan Tompson, Yann LeCun, and Christoph Bregler. MoDeep: A deep learning framework using motion features for human pose estimation. In *Proceedings of the Asian Conference on Computer Vision*, 2015.

[23] Jonathan Tompson, Arjun Jain, Yann LeCun, and Christoph Bregler. Joint training of a convolutional network and a graphical model for human pose estimation. In *Proceedings of the Advances in Neural Information Processing Systems (NIPS)*, 2014.

[24] Jonathan Tompson, Ross Goroshin, Arjun Jain, Yann Lecun, and Christoph Bregler. Efficient object localization using convolutional networks. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

[25] Alexander Toshev and Christian Szegedy. DeepPose: Human pose estimation via deep neural networks. *Computing Research Repository (CoRR)*, 2013.

[26] James Charles, Tomas Pfister, Derek Magee, David Hogg, and Andrew Zisserman. Personalizing human video pose estimation. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, nov 2016.

[27] Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. Realtime multi-person 2D pose estimation using part affinity fields. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, nov 2017.

[28] Felix Achilles, Alexandru Eugen Ichim, Huseyin Coskun, Federico Tombari, Soheyl Noachtar, and Nassir Navab. Patient MoCap: Human pose estimation under blanket occlusion for hospital monitoring

applications. In *Proceedings of International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, 2016.

[29] Shuangjun Liu, Yu Yin, and Sarah Ostadabbas. In-bed pose estimation: Deep learning with shallow dataset. In *arXiv Preprint*, nov 2017.

[30] Vasileios Belagiannis, Xinchao Wang, Horesh Ben Shitrit, Kiyoshi Hashimoto, Ralf Stauder, Yoshimitsu Aoki, Michael Kranzfelder, Armin Schneider, Pascal Fua, Slobodan Ilic, Hubertus Feussner, and Nassir Navab. Parsing human skeletons in an operating room. *Machine Vision and Applications*, 27, 2016.

[31] Abdolrahim Kadkhodamohammadi, Afshin Gangi, Michel de Mathelin, and Nicolas Padoy. A multi-view RGB-D approach for human pose estimation in operating rooms. In *Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV)*, jan 2017.

[32] Alexandros Chaaraoui, José Padilla-López, Francisco Ferrández-Pastor, Mario Nieto-Hidalgo, and Francisco Flórez-Revuelta. A vision-based system for intelligent monitoring: Human behaviour analysis and privacy by context. *Sensors*, 14(5):8895–8925, 2014.

[33] Francis Seung-hyun Baek. Autonomous patient safety assessment from depth camera based video analysis. *Dissertation at UC San Diego*, 2016.

[34] Karel Zuiderveld. Contrast limited adaptive histogram equalization. In *Graphics Gems*, pages 474–485. 1994.

[35] Stephen M. Pizer, E. Philip Amburn, John D. Austin, Robert Cromartie, Ari Geselowitz, Trey Greer, Bart Ter Haar Romeny, John B. Zimmerman, and Karel Zuiderveld. Adaptive histogram equalization and its variations. *Computer Vision, Graphics, and Image Processing*, 39:355–368, 1987.

[36] Aouaidjia Kamel, Bin Sheng, Po Yang, Ping Li, Ruimin Shen, and David Dagan Feng. Deep convolutional neural networks for human action recognition using depth maps and postures. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, (99), 2018.

[37] Kezhi Li, John Daniels, Pau Herrero-viñas, Chengyuan Liu, and Pantelis Georgiou. Convolutional recurrent neural networks for blood glucose prediction. *arXiv Preprint*, 2018.

[38] Tom Young, Devamanyu Hazarika, Soujanya Poria, and Erik Cambria. Recent Trends in Deep Learning Based Natural Language Processing, 2017.

[39] Steve Lawrence, C. Lee Giles, Ah Chung Tsoi, and Andrew D. Back. Face recognition: A convolutional neural-network approach. *IEEE Transactions on Neural Networks*, 1997.

[40] Nal Kalchbrenner, Edward Grefenstette, and Phil Blunsom. A convolutional neural network for modelling sentences. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2014.

[41] Daojian Zeng, Kang Liu, Siwei Lai, Guangyou Zhou, and Jun Zhao. Relation classification via convolutional deep neural network. *Coling*, 2014.

[42] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017.

[43] James Charles, Tomas Pfister, Mark Everingham, and Andrew Zisserman. Automatic and efficient human pose estimation for sign language videos. *International Journal of Computer Vision*, 110:70–90, 2014.

[44] Tsung Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft COCO: Common objects in context. In *Proceedings of the European Conference on Computer Vision (ECCV)*, may 2014.

[45] Ben Sapp and Ben Taskar. MODEC: Multimodal decomposable models for human pose estimation. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013.

[46] Mykhaylo Andriluka, Leonid Pishchulin, Peter Gehler, and Bernt Schiele. 2D human pose estimation: New benchmark and state of the art analysis. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.

[47] Gunnar Farnebäck. Two-frame motion estimation based on polynomial expansion. In *Proceedings of the 13th Scandinavian Conference on Image Analysis (SCIA)*, 2003.

[48] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. Caffe: Convolutional architecture for fast feature embedding. In *Proceedings of the 22nd ACM International Conference on Multimedia*, 2014.

[49] Rudolf E. Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82:35, 1960.

[50] Aleksandr Aravkin, James V. Burke, Lennart Ljung, Aurelie Lozano, and Gianluigi Pillonetto. Generalized Kalman smoothing: Modeling and algorithms. *Automatica*, 86:63–86, 2017.

[51] Byron M Yu and Krishna V Shenoy. Derivation of Kalman filtering and smoothing equations. Technical report, Stanford University, 2004.

[52] Ramsey Faragher. Understanding the basis of the Kalman filter via a simple and intuitive derivation. *IEEE Signal Processing Magazine*, 29:128–132, 2012.

[53] Greg Welch and Gary Bishop. An introduction to the Kalman filter. In *Proceedings of the ACM Special Interest Group on Computer Graphics (SIGGRAPH)*, 2001.

[54] X. Rong Li and Vesselin P. Jilkov. Survey of maneuvering target tracking. Part I: Dynamic models. *IEEE Transactions on Aerospace and Electronic Systems*, 2003.

[55] Pieter Abbeel, Adam Coates, Michael Montemerlo, Andrew Y. Ng, and Sebastian Thrun. Discriminative training of Kalman filters. In *Proceedings of Robotics: Science and Systems I*, 2005.

[56] Laurens Van Der Maaten and Geoffrey Hinton. Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9:2579–2605, 2008.

Chapter 6

Big Picture Discussion

6.1 Summary of Contributions

The works presented in Chapters 2–5 describe methods to collect and annotate unstructured behaviors from the epilepsy monitoring unit, as well as analyses to functionally map and classify neural activity with respect to behaviors labeled in a coarse manner.

Chapter 2 was our preliminary demonstration of the continuous video and neural recording system used throughout the entirety of the presented study. As a proof-of-concept, we showed that neural activity from the continuous monitoring paradigm can be used to map electrode response to left- and right-sided movements tracked using optical flow of an edge-based hand detector.

Chapter 3 evaluated whether continuously recorded intracortical signals could be used to classify coarsely labeled behavior contexts. We demonstrated that traditional features (high-frequency band spectral power) could be used to distinguish between abstract behavior contexts. We also extended our electrode assessment and functional mapping methods to the sEEG/depth modality.

Chapter 4 demonstrated an evaluation framework to study neural activity related to natural movements not evoked by a task, annotated over hours of video. We found that neural signal features previously identified from task-based experiments are also modulated before and during a variety of movement behaviors. These movement behaviors were coarsely labeled by time period and movement side using dense optical flow, meaning movements within a label represented a wide variety of uninstructed behaviors. A rigorous nested cross-validation framework was used to classify both movement onset and lateralization with statistical significance for three subjects.

Finally, Chapter 5 attempted to improve the fidelity of movement annotations extracted from clinical video by way of a semi-automated pose tracking framework. The extended system trained subject-specific convolutional neural network models on a subset of patient-specific video recordings chosen to maximize

the variety of poses trained on. Scene lighting changes were also accommodated. Afterwards, these estimations were refined over time through a Kalman filter with fitted noise parameters.

All in all, these collective works address the general challenges of recording, annotating, and evaluating neural correlates to unstructured human behaviors. The preliminary nature of this work means that a majority of our demonstrations are whether the continuous paradigm can be leveraged, how one might go about leveraging it, and evaluations that tie our results back to earlier task-based studies. The context of this work is specifically the epilepsy monitoring unit. Our advances here motivate future works that focus more intently on what types of behaviors and neural signal features to explore.

6.2 Lessons Learned

We learned many things over the course of this study, which led to discussions both published and spoken. These lessons are consolidated here, while further details can be found in each previous chapter.

6.2.1 Continuous and Passive Monitoring in the Clinic

Collecting data in clinical environments is challenging for multiple reasons. Before a new recording system can deployed, it has to go through several designs and reviews as assessed by multiple fronts. This process can add several months to a year to a new study, but is critical for human-focused research.

On the patient side, our monitoring system was designed to be as minimally intrusive during the study. This meant that it could not produce any irritating noises, sounds, or any other offenses to their senses. Had the system included wearable sensors, for example, we would have to show that wearing said sensors for a long period of time would not cause skin irritations. Protecting patient privacy was also important, both during data collection and subsequent offline analysis. As part of the study protocol, for example, we provided a log that the participants and caregivers could use in order to censor certain times of the day. Whether there was a need for biological cycle or if a delicate matter was being discussed, logged periods were immediately deleted. These recorded data were also anonymized prior to analysis, which really means that videos were processed into non-identifying features before anyone saw them.

On the clinical side, our monitoring system had to be mobile with a small footprint, so that the workflow of nurses and physicians would not be impeded. This also meant familiarizing clinicians, research coordinators, and nurses with the system. The clinical environment introduces a lot of confounding noise due to the amount of motion and equipment around the subjects. Factors like outside movement, scene lighting changes, and external electromagnetic noise had to be accounted for in offline analyses, as discussed

in Chapters 2–5.

On the system side, our monitoring system was built to collect consistent and synchronized signals. Ensuring the quality of data collected and establishing tests to routinely monitor the status of data quality turned out to be incredibly important endeavors at the beginning of this study. This also extended to the synchronizing hardware we built, which was designed to align our video and neural signals as closely as possible (at the frame level).

IRB Protocols

As part of human-focused research, a protocol must be submitted to and approved by the Institutional Review Board (IRB) of each institution in which the study is being conducted. The protocol consists of a formal set of documents detailing the study, specifically focusing on subject protection, as described in the previous section. Between writing the protocol and several types of consent forms, collecting letters of support from all teams involved in the study, and waiting on/addressing feedback from the IRB, expect six to nine months before the study is formally approved.

6.2.2 Nuances of Unstructured Behaviors

We found that when patient behavior was monitored completely passively, in that none of their movements or behaviors were instructed by the researcher, the total number of observed active behaviors was low. Even when labeling broad movement periods using optical flow, the ratio of non-moving to moving periods (that were relevant) was almost 100:1. Clinically this was to be expected, as the study participants were recovering from neurosurgery, under the influence of various medications, and prone to seizures. In addition, many of the movement periods excluded from study were confounded or perturbed by external activity. For example, mealtimes were typically excluded even when the subject was moving because they were both obscured by their bedside table as well as movement from their caregivers.

The annotation methods in this work applied to either individual video frames or segments of video made of sequences of frames. These manual and semi-automated methods could a) mark on/offset timestamps of behaviors defined by a rubric; b) quantize the position of joints representing subject posture in each frame; c) calculate optical flow between sequential pairs of frames. These annotations are intended to represent the movement behaviors being studied, but limiting in what exactly the represented. On one hand, the coarse movement labels from Chapters 3 and 4 limited the number of distinct classes of behaviors being evaluated to an appropriate level given the sample size. In doing so, however, we disregarded the nuances between, say, the variety of left-sided movements that were possible. On the other hand, the raw values of

pose estimation and optical flow from Chapters 4 and 5 produced data consisting of many extracted values agnostic to the actual behaviors. For example, this meant that additional parsing was required to turn a sequence of x,y positions of the right hand to some other representation conducive to rapid analysis. With the potential to observe a wide variety of behaviors in the unstructured paradigm, new ways to describe and relate these behaviors are clearly necessary.

6.2.3 Neural Signal Processing

Our neural feature selection was primarily motivated by previous task-based ECoG analyses like [1] and [2]. These works identified spectral responses to various movement modalities in the high-frequency band and low-frequency band of cortical signals near central sulcus. These frequency bands were fixed at 76–100Hz and 8–32Hz, respectively, though these values tend to vary across other works. This standard approach also averages the filtered signals over a fixed window time, as a measure of signal power. While these methods are drawn from the field of digital signal processing, they are inconsiderate of the underlying properties of electrophysiological signals, in that they exhibit both periodic and aperiodic features. Recent evidence shows that these aperiodic features can be dynamic with factors like age, task demands, and cognitive state. Methods like [3] and [4] can enable informed analyses of neural correlates to unstructured behaviors, which we hope to motivate into future work.

In a related direction, there are works like [5], [6], and [7] exploring methods to learn informative encodings from raw neural signals. A particular strength of these methods is their ability to compress the increasing number of neural channels being recorded into informative encodings. Strategies that scale with increasingly larger data sets, both in terms of number of sensors and subjects, may allow for analysis of individual behaviors and their underlying neural activity.

6.3 Proof-of-Concept Results

6.3.1 Movement Sequences in 3D

While computer vision algorithms to track pose from video (which is generally 2D) are continuing to improve, we are still far from having consistently confident and reliable methods to track pose in 3D without wearing some suite of sensors such as a motion capture suit covered in infrared (IR) markers. Using our depth and infrared recording modalities and human annotation, we began generating example datasets of segmented movement sequences. To generate these 3D data, we mapped manual annotates from the infrared images onto their corresponding depth frames. The pixel x and y locations, combined with the

depth reading, can be transformed into real-world coordinate systems and is centered around the sensor by default. With this added dimension, these pose sequences are more representative of the subject's movement as compared to their 2D counterpart in Chapter 5.

6.3.2 Characterization of Individual Movement Sequences

To begin addressing, or rather organizing, our 3D movement sequences, we drew from methods in the machine learning literature focused on encoding sequential data of variable lengths [8]. Discovering sequence clusters can reveal latent characteristics of subject movement and can also support high-level applications such as neural signal analysis, clinical motor evaluations, and anomaly detection. This work addresses the movement sequence clustering problem by developing a method that can detect space- and time- invariant movement clusters. Given the input sequences, our goal was to convert each trajectory into a fixed-length representation that well encodes the subject's movement behavior. Once these high-quality trajectories are learned, we can then apply clustering algorithms on the fixed-length representations. We then show that there is mutual information between movement kinematics and these representations. We demonstrated this framework by re-characterizing monkey reaching movements from the [9] center-out reach task.

6.4 Future Directions

These contributions to the study of neural correlates to unstructured motor behaviors are but one step towards practical destinations like passive functional mapping of and generalizing neuroprosthetics to naturalistic behaviors. Our advances here motivate future works that focus more intently on what types of behaviors and neural signal features to explore. In the short term, these directions aim to analyze large-scale multi-modal datasets of unstructured motor behaviors from the epilepsy monitoring unit. Cortical coverage in this setting is entirely driven by patient need, which means it may not be possible to study nuances in behavior just yet. However, the broad coverage represented in such datasets may allow us to question broader areas such as the connectivity between cortical regions and how they relate to spatio-temporal patterns of neural signal indicative of abstract brain states. In the long term, we hope our demonstrations motivate new iterations of the unstructured experiment paradigm. Our works can serve as starting point for designing, deploying, and analyzing experiments in clinical settings. The need for methods that monitor and annotate multi-modal clinical data remain important. Accepting that our ability to interface with the brain and behavior will continue to evolve with growing technology, it is difficult to predict exactly how

future experiments will be designed. Some of the challenges these experimenters will face may be similar to the ones described in these works. If you are one such researcher, we hope this writing helped you in some way. Good Luck.

References

[1] Kai J Miller, Eric C Leuthardt, Gerwin Schalk, Rajesh PN Rao, Nicholas R Anderson, Daniel W Moran, John W Miller, and Jeffrey G Ojemann. Spectral changes in cortical surface potentials during motor movement. *Journal of Neuroscience*, 27(9):2424–2432, 2007.

[2] Wei Wang, Alan D Degenhart, Jennifer L Collinger, Ramana Vinjamuri, Gustavo P Sudre, P David Adelson, Deborah L Holder, Eric C Leuthardt, Daniel W Moran, Michael L Boninger, et al. Human motor cortical activity recorded with micro-ecog electrodes, during individual finger movements. In *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 586–589. IEEE, 2009.

[3] Matar Haller, Thomas Donoghue, Erik Peterson, Paroma Varma, Priyadarshini Sebastian, Richard Gao, Torben Noto, Robert T Knight, Avgusta Shestyuk, and Bradley Voytek. Parameterizing neural power spectra. *bioRxiv*, page 299859, 2018.

[4] S Cole, T Donoghue, R Gao, and B Voytek. Neurodsp: A package for neural digital signal processing. *Journal of Open Source Software*, 4(36), 2019.

[5] Andrzej Cichocki. Tensor decompositions: a new concept in brain data analysis? *arXiv preprint arXiv:1305.0395*, 2013.

[6] David Sussillo, Rafal Jozefowicz, LF Abbott, and Chethan Pandarinath. Lfads-latent factor analysis via dynamical systems. *arXiv preprint arXiv:1608.06315*, 2016.

[7] Tejaswy Pailla, Kai Joshua Miller, and Vikash Gilja. Autoencoders for learning template spectrograms in electrocorticographic signals. *Journal of neural engineering*, 2018.

[8] Di Yao, Chao Zhang, Zhihua Zhu, Jianhui Huang, and Jingping Bi. Trajectory clustering via deep representation learning. In *2017 international joint conference on neural networks (IJCNN)*, pages 3880–3887. IEEE, 2017.

[9] Pavan Ramkumar, Brian Dekleva, Sam Cooler, Lee Miller, and Konrad Kording. Premotor and motor cortices encode reward. *PloS one*, 11(8):e0160851, 2016.